



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Schweifer *et al.*

Appl. No. 09/899,569

Filed: July 6, 2001

For: **Tumorassoziiertes Antigen (B345)**

Confirmation No. 1574

Art Unit: *To Be Assigned*

Examiner: *To Be Assigned*

Atty. Docket: 0652.2280001/EKS/AES

**Submission of Certified Copy of 35 U.S.C. § 119(a)-(d)
Priority Documents In Utility Application**

Commissioner for Patents
Washington, D.C. 20231

Sir:

Submitted herewith are two (2) certified copies of Applicants' U.S.C. § 119(a)-(d) priority documents, to perfect the claim to priority filed on July 6, 2001.

Country	Priority Document Appl. No.	Filing Date
Germany	100 33 080.0	July 7, 2000
Germany	101 19 294.0	April 19, 2001

Prompt acknowledgment of this submission is respectfully requested.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

A handwritten signature in cursive ink that appears to read "Ann E. Summerfield".

Ann E. Summerfield
Agent for Applicants
Registration No. 47,982

Date: February 26, 2002

1100 New York Avenue, N.W.

Suite 600

Washington, D.C. 20005-3934

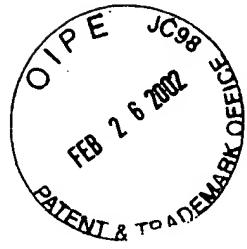
(202) 371-2600

P:\USERS\Ann\Angela\0652\0652.2280001\SubofPriorDoc

BUNDESREPUBLIK DEUTSCHLAND



CERTIFIED COPY OF
PRIORITY DOCUMENT



Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen:

101 19 294.0

Anmeldetag:

19. April 2001

Anmelder/Inhaber:

Boehringer Ingelheim International GmbH,
Ingelheim/DE

Bezeichnung:

Tumorassoziiertes Antigen (B345)

IPC:

C 07 K, C 12 N

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der
ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 30. Juli 2001
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Levitsky

Case 12/219 DI Fa/dc

BOEHRINGER INGELHEIM INTERNATIONAL GMBH
55216 Ingelheim am Rhein (BRD)

Tumorassoziiertes Antigen (B345)

Zusammenfassung

5

Tumorassoziiertes Antigen B345 und dafür kodierende
DNA-Moleküle.

10

Die Erfindung bezieht sich auf die Chemotherapie von Tumorerkrankungen.

Normale Körperzellen unterliegen einem strikt geordneten System, das das Wachstum, die Zellteilung und das Absterben bestimmter Zellen kontrolliert. So teilen sich Körperzellen einer erwachsenen Person nur dann, wenn sie tote Zellen ersetzen oder eine Verletzung verheilen müssen. Krebszellen dagegen wachsen unkontrolliert weiter, sie akkumulieren und bilden einen Tumor.

Erreicht der Tumor eine kritische Größe, können Krebszellen über die Blutbahnen oder das Lymphsystem auch in andere Bereiche des Körpers transportiert werden und dort Kolonien bilden (Metastasen). Nicht alle Tumore sind kanzerogen, denn benigne Tumore metastasieren nicht und sind daher meistens nicht lebensgefährlich, da sie chirurgisch entfernt werden können. Detailliertere Informationen zu diesem Thema sowie den weiter unten diskutierten Aspekten der Tumorentstehung geben folgende Publikationen: Rauscher und Vogt, 1997; Kastan, 1997; Hesketh, 1995; Puszta, Lewis und Yap, 1995.

Die Transformation einer gesunden Zelle in eine Krebszelle kann durch eine ganze Reihe von Faktoren, wie Umwelteinflüsse, Strahlung, Viren oder chemische Reagenzien ausgelöst werden. Bei der Tumorentwicklung spielen jedoch auch epigenetische (Methylierungen, Acetylierungen und veränderte Chromatinstruktur) und genetische Modifikationen (Punktmutation, Deletion, Amplifikation, Translokation) eine bedeutende Rolle.

Mutationen in kodierenden Bereichen von Genen, die an der Regulation der Zell-Proliferation beteiligt sind, können an der Überführung einer normalen Zelle in eine Tumorzelle beitragen, da die transformierte Zelle

5 Wachstumsvorteile gegenüber ihrer gesunden Nachbarzelle hat.

Krebs entsteht also durch eine Ansammlung von vererbten oder erworbenen Mutationen in kritischen Protoonkogenen oder Tumorsuppressorgenen.

10 Die Zellproliferation steht unter der Kontrolle verschiedener Gensysteme, während Produkte von Onkogenen an der Signalvermittlung von Zelloberfläche zum Zellkern beteiligt sind, geleiten cyclinabhängige Proteinkinase und ihre Inhibitoren die Zelle durch den Zellzyklus.

15 Nicht selten werden Störungen in der Synthese dieser Proteine bei Tumorzellen gefunden. Eine zentrale Rolle spielt dabei das p53-Protein.

Proteine vom Typ des RB-Proteins regulieren die Verfügbarkeit entscheidender Transkriptionsfaktoren.

20 Die in Tumorgeweben hochregulierten Gene stellen meist den Ausgangspunkt für weitere detaillierte Analysen dar und da Proteine verschiedenster Funktionen hochexprimiert werden, ist der Ansatz für therapeutische Interventionen sehr vielseitig. Es ist daher, das Ziel 25 der Krebsforschung weitere neue Zielmoleküle' (sog. „Targets“) für therapeutische Interventionen zu finden, die dann für eine gezielte Therapie mit geringen Nebenwirkungen benutzt werden können.

In erster Linie gilt es daher molekulare Veränderungen zwischen Normalgewebe und Tumor auf dem Niveau der Genexpression („Transkriptionslevel“) aufzudecken, die einerseits neue Targets identifizieren sollen und

5 andererseits für die Entwicklung bzw. das Auffinden von Substanzen zur Hemmung von Fehlfunktionen herangezogen werden können.

Eine ganze Reihe verschiedenster Methoden zur Identifizierung und Charakterisierung von neuen Targets,

10 die den Ausgangspunkt für die Entwicklung von neuen Therapeutika darstellen, basieren auf der Erstellung differenzieller mRNA Transkriptionsprofile zwischen Tumoren und Normalgeweben. Dazu zählen die differenzielle Hybridisierung, die Erstellung von
15 Subtraktions-cDNA-Banken ("representational difference analysis"; Hubank und Schatz, 1994; Diatchenko et al., 1996) und der Einsatz der DNA-Chip Technologie oder der SAGE-Methode (Velculescu et al., 1995).

Neben immuntherapeutischen Ansätzen kommt der gezielten

20 Chemotherapie eine wesentliche Rolle bei der Behandlung von Krebs zu. Unter Chemotherapie versteht man die Verabreichung von Substanzen, die durch Eingriff in Stoffwechsel, Signaltransduktion und Zellteilungsvorgänge maligner Zellen entweder
25 zytostatisch oder zytotoxisch-zytolytisch wirken.

Chemotherapeutika lassen sich auf Grund der Beeinflussung von spezifischen Targets in der Tumorzelle, nach Art der zellulären Interaktion und der Wechselwirkung mit einer bestimmten Zellzyklusphase, in
30 verschiedene Kategorien unterteilen.

Die Art der Krebsbehandlung hängt vom Tumorstadium ab, entscheidend ist dabei, ob bereits Metastasen vorhanden sind und wie weit diese im Körper verbreitet sind. Die Applikation von Zellgiften zur Krebsbehandlung ist,

5 neben operativen Maßnahmen und der Strahlentherapie, ein integraler Bestandteil der heutigen Therapiekonzepte in der Onkologie.

Im wesentlichen gibt es zwei Hauptziele der Chemotherapie: In erster Linie ist es die Heilung von

10 Krebs; das bedeutet, dass der Tumor verschwindet und nicht mehr auftritt. Ist eine Heilung aus verschiedenen Gründen nicht mehr möglich, versucht man den Tumor in seinem Wachstum und seiner Ausbreitung einzuschränken bzw. zu kontrollieren.

15 Prinzipiell entfalten Substanzen, die bei der Chemotherapie Anwendung finden, ihre Wirkung bei allen sich teilenden Zellen. Tumorzellen zeigen jedoch eine höhere Empfindlichkeit gegenüber Chemotherapeutika als gesunde Zellen, da hauptsächlich stark proliferierende
20 Zellen angegriffen werden.

Jedes Gewebe hat ein charakteristisches Wachstumsverhalten, das Zellteilung, Wachstumsstillstand, Differenzierung und Alterung umfasst und durch interne und externe Faktoren beeinflusst und reguliert wird.

25 Viele der heute eingesetzten zytotoxischen Chemotherapeutika wirken nur bei proliferierenden Zellen (nicht in der G0-Phase der Zellteilung). Dabei werden allerdings sowohl Normal- als auch Krebszellen attackiert. Die Zerstörung von normalen Zellen kann zu
30 starken Nebeneffekten führen; z.B. Zerstörung der

Blutzellen-produzierenden Gewebe des Knochenmarks
(Myelosuppression).

Chemotherapeutika werden je nachdem, wie sie spezifische Substanzen innerhalb der Tumorzelle beeinflussen, mit

5 welchen zellulären Prozessen die Medikamente interagieren und welche Zellzyklusphase sie beeinflussen, in verschiedene Kategorien unterteilt.
Diese Informationen sind für Onkologen notwendig für die Entscheidung welche Präparate bei der Therapie
10 miteinander kombiniert werden können.

Die in Tumorgeweben hochregulierten Gene stellen somit potentielle neue Zielstrukturen dar und da Proteine verschiedenster Funktionen hochexprimiert werden, ist der Ansatz für therapeutische Interventionen sehr
15 vielseitig.

Es ist daher das Ziel der Krebsforschung, weitere neue Targets für therapeutische Interventionen zu finden, die dann für eine gezielte Therapie mit - verglichen mit derzeitig eingesetzten Therapeutika - geringeren

20 Nebenwirkungen benutzt werden können.

In Tumorgeweben hochregulierte Gene stellen Angriffspunkte und somit potentielle Zielstrukturen („Targets“) für die Chemotherapie dar.

Aufgabe der vorliegenden Erfindung war es, ein neues,
25 bevorzugt von Tumorzellen exprimierte Protein bereitzustellen, das ein Zielmolekül für die Intervention mittels chemotherapeutischer Methoden darstellt.

Diese Aufgabe wurde gelöst, indem zunächst mittels RDA („representational difference analysis“) zwischen einer Lungen-Adenokarzinom-Zelllinie (A549) und Normallungengewebe eine cDNA-Substraktionsbibliothek hergestellt wurde. Zur Selektion der im Tumor überexprimierten Antigene wurden anschließend die erhaltenen cDNA-Klone sequenziert und mit in Datenbanken verfügbaren Sequenzen verglichen. Unter den dabei annotierten Genen befanden sich 321 unbekannte, zu denen größtenteils ESTs („expressed sequence tags“) -Einträge in der Datenbank existierten. Nach einer weiteren qualitativen PCR-Analyse in cDNA-Bibliotheken von kritischen Normalgeweben und immunprivilegierten Geweben sowie detaillierteren Datenbankrecherchen wurde die Anzahl der Kandidatenklone auf 59 eingeschränkt, deren ESTs nicht aus kritischen Normalgeweben stammen.

Diese Klone wurden auf Incyte DNA Chips gespottet und mit einer ganzen Reihe von Tumorgeweben und Normalgeweben als Referenz hybridisiert. Das mRNA Expressionsprofil von EST Fragmenten, die in Krebsgeweben und Normalgeweben differentiell exprimiert werden und zu einem noch unbekannten Gen gehören, wurde mit unterschiedlichen Methoden verifiziert.

Die Länge der Transkripte wurde mittels Northern blot Analyse bestimmt und das Expressionsmuster in verschiedenen Zellsystemen durch quantitative PCR exakt charakterisiert. Nur unbekannte Gene bzw. ESTs mit tumorspezifischen Expressionsprofil wurden weiterverfolgt und einer "full length Klonierung" unterworfen. Potentielle ORFs („open reading frames“) werden in die entsprechende Aminosäuresequenz

umgewandelt und zur möglichen Funktionsvorhersage mittels *in silico* Strategien analysiert.

Die humane B345-cDNA wurde kloniert, die in einem ersten Klonierungsansatz erhaltene Sequenz ist in SEQ ID NO:1 dargestellt. Die Sequenzanalyse der in diesem Ansatz klonierten humanen B345-cDNA zeigte einen durchgehenden offenen Leserahmen von Position 215 bis Position 2461 (exklusive Stopcodon), der, auf Nukleotid- und Proteinebene, in den bekannten Sequenzen der Datenbanken keine Homologie oder Identität aufweist. Aus den aus Northern Blot Experimenten gewonnenen Daten ist zu schließen, dass das B345-Transkript eine Länge von ca. 6,5 kb hat. In einem ersten Ansatz wurde als klonierter Bereich eine B345-cDNA mit 5897 bp (exklusive polyA-Region), erhalten, wobei das Vorhandensein eines Polyadenylationssignals und des PolyA-Tails am 3'-Ende der Sequenz auf die Vollständigkeit der cDNA in diesem Bereich hindeutete. Aufgrund der Tatsache, dass im 5'-Bereich der klonierten cDNA von Position 1 bis 214 kein durchgehender Leserahmen aufschien, wurde zunächst angenommen, dass es sich bei dem ATG an Position 215, die auch zu 75% einer Kozak Translationinitiationsstelle (ACCATGT) (Kozak, 1987) entspricht, um das Startkodon von B345 handelt.

In einem weiteren Klonierungsansatz wurden mittels einer molekularbiologischen Standardmethode, und zwar mittels sog. „Promotor Finder DNA Walking“, zusätzliche Informationen über die weiter stromaufwärts liegende Sequenz von B345 gewonnen.

Somit wurde die in dem ersten Klonierungsversuch
erhaltene B345-Sequenz (SEQ ID NO:1) in der 5`Region
erweitert. Der Transkriptionsstart konnte unter
Anwendung der Primer Extension Analyse genau lokalisiert
5 werden und liegt bei Position 201 (SEQ ID NO:3). Durch
wiederholte Sequenzierungen auch in der 3`Region wurde,
im Vergleich zur in SEQ ID NO:1 dargestellten Sequenz,
eine zusätzliche Base an Position 2430 gefunden, die zu
einer Leserasterverschiebung führt und somit das
10 Stopcodon von Position 2729 auf 2791 verschiebt. Die
erhaltene cDNA (SEQ ID NO:3) besitzt einen offenen
Leserahmen, der für ein potentielles Protein mit einer
Länge von 836 Aminosäuren (SEQ ID NO:4) kodiert. Die
Translationsinitiationsstelle an Position 283 entspricht
15 etwa zu 70% einer Kozak-Konsensussequenz.

Die Promotorregion 200bp upstream der mutmaßlichen
Transkriptionsstartstelle enthält weder eine TATA noch
eine CCAAT box, jedoch eine eindeutige GC-box, welche
eine Bindungsstelle des SP1 Proteins darstellt. Die
20 Tatsache, dass der GC Gehalt in der 5`Region über 60%
ist, deutet auf ein CpG Island hin (Bird, 1986).

Die resultierende primäre Aminosäuresequenz von B345 ist
in SEQ ID NO:4 gezeigt. Die Analyse des Hydrophilizitäts
Plots der Aminosäuresequenz zeigt, dass das B345 Protein
25 zwei charakteristische hydrophobe Domänen aufweist, die
ein Signalpeptid und eine helikale Transmembrandomäne
darstellen (Fig. 6). Diese polarisierte Struktur deutet
darauf hin, dass es sich bei B345 um ein integrales
Membranprotein handelt.

Die extrazelluläre Domäne lässt auf die Existenz von definitiv einer, gegebenenfalls drei, CUB Domänen schließen. CUB Domänen kommen bei verschiedenen, meist während der Embryonalentwicklung regulierten Proteinen vor. Eine kürzlich erschienene Publikation (Gerstein et al., 2000) demonstriert, daß CUB Domänen enthaltende Proteine die am ausgeprägtesten differenziell regulierten Proteine in *C. elegans* sind. Da Gene, die eine Schlüsselrolle in der Embryonalentwicklung haben, entsprechende Funktionen, z.B. in der Zellteilung, der Zellproliferation oder Signalübertragung, in Krebs ausführen, kann angenommen werden, dass eine Überexpression von B345 in Zellen eine Veränderung in den Eigenschaften der Substratadhäsion oder der extrazellulären Matrix bewirkt. Das B345 Protein weist 12 potentielle N-Glycosylierungsstellen auf, die in der mutmaßlichen extrazellulären Domäne zu finden sind.

Aufgrund seiner Aminosäuresequenz ist anzunehmen, dass das B345-Protein eine β -Sheet Sekundärstruktur ausbildet, da sich CUB Domänen bekanntlich als β -Sandwich falten.

Die intrazelluläre Domäne (Bereich 691-836) weist keine signifikanten Homologien auf. Der gesamte C-Terminus zeigt jedoch eine Identität (82%) über 124 Aminosäuren mit einem EST (Acc No. AW063026) aus humanen Eierstockkrebs Zellen.

Ausgehend von den Funktionen anderer CUB Domänen enthaltender Proteine kann gefolgert werden, daß das B345 Transmembranprotein in der Kommunikation, der Interaktion und/oder der Signaltransduktion mit

extrazellulären Komponenten oder Liganden eine Rolle spielt. Ferner sind die Daten der Expressionsanalyse ein starkes Indiz dafür, dass B345 beim metastatischen Prozess von Krebs, insbesondere Dickdarmkrebs, beteiligt 5 ist.

Für die Aufklärung der physiologischen Funktion und der Rolle von B345 bei der Metastasierung sind folgende Untersuchungsmethoden geeignet:

Zunächst werden Zelllinien, vorzugsweise humane

10 Zelllinien identifiziert, z.B. mittels TaqMan PCR, die B345 nicht endogen exprimieren. Die Zellen werden mit einem Plasmid, das die B345-Sequenz enthält, transfiziert und B345 exprimiert. Änderungen in der Morphologie und/oder dem Migrationsverhalten, das z.B. 15 mittels Softagar Assay (Hamburger und Salmon, 1977) oder Migrationsassay (Liaw et al; 1995) der B345 exprimierenden Zellen gegenüber den nicht-transfizierten Zellen deuten auf eine Rolle von B345 in dem dafür 20 verantwortlichen biologischen Prozess hin. Dies ist ein deutlicher Hinweis auf eine Beteiligung von B345 an der Interaktion von Tumorzellen untereinander und/oder mit der extrazellulären Matrix und somit auf eine Funktion bei der Metastasierung.

Alternativ bzw. zusätzlich zu dieser Funktionsanalyse

25 wird in einem komplementären Ansatz die Expression von B345 in Zellen, die dieses Protein endogen exprimieren, unterdrückt, um ebenfalls die etwaige Änderungen in Morphologie und/oder Migrationsverhalten festzustellen.

Außerdem wird gegebenenfalls untersucht, ob

30 Proteinkomponenten existieren, die mit B345 inter- oder

extrazellulär interagieren (z.B. mittels Yeast Two Hybrid System (Fields und Song, 1989).

Die Erfindung betrifft somit in einem ersten Aspekt ein tumorspezifisches Polypeptid mit der Bezeichnung B345, 5 mit der in SEQ ID NO: 4 angegebenen Aminosäuresequenz oder ein Polypeptid, das von einem Polynukleotid kodiert wird, das unter stringenten Bedingungen mit einem Polynukleotid der in SEQ ID NO:3 dargestellten Sequenz oder einer Teilsequenz davon hybridisiert, sowie davon 10 abgeleitete Proteinfragmente oder Peptide.

In einem weiteren Aspekt betrifft die vorliegende Erfindung ein isoliertes DNA-Molekül, kodierend für das tumorspezifische Polypeptid der Bezeichnung B345.

Bevorzugt ist das erfindungsgemäße DNA-Molekül ein 15 Polynukleotid der in SEQ ID NO: 3 dargestellten Sequenz oder ein Fragment davon oder ein DNA-Molekül, das mit einem DNA-Molekül der in SEQ ID NO: 3 dargestellten Sequenz oder einer Teilsequenz davon unter stringenten Bedingungen hybridisiert, oder ein Fragment davon.

20 Unter „stringenten Bedingungen“ wird z.B. verstanden: Inkubation über Nacht bei 65°C - 68°C mit 6xSSC (1xSSC = 150 mM NaCl, 15 mM Tri-Natriumcitrat), 5xDenhardt's Lösung, 0.2%SDS, 50 µg/ml Lachsspermien-DNA, daran anschließend Waschen zweimal 30 min mit 2xSSC, 0.1% SDS 25 bei 65°C, einmal 30 min mit 0.2xSSC, 0.1% SDS bei 65°C und gegebenenfalls abschließendes Spülen mit 0.1xSSC, 0.1%SDS bei 65°C, oder äquivalente Bedingungen.

Die erfindungsgemäßen DNA Moleküle kodieren für (Poly)peptide der Bezeichnung B345 mit der in SEQ ID NO:

4 dargestellten Aminosäuresequenz bzw. für davon
abgeleitete Proteinfragmente oder Peptide; damit sind
DNA Moleküle bzw. Fragmente mitumfasst, die durch die
Degeneration des genetischen Codes Abweichungen von der
5 in SEQ ID NO: 3 dargestellten Sequenz aufweisen

In einer Ausführungsform der Erfindung betrifft die
Erfindung ein isoliertes DNA-Molekül der in SEQ ID NO:3
dargestellten Sequenz oder ein Fragment davon, oder ein
DNA-Molekül, das mit einem DNA-Molekül der in SEQ ID
10 NO:3 dargestellten oder mit einer Teilsequenz davon
hybridisiert, kodierend für das natürliche
B345-Polypeptid bzw. für ein Fragment davon.

Die B345-DNA-Moleküle können in einer sog. DNA-Vakzine
für die Immuntherapie von Tumoren verwendet werden.

15 Dabei können die B345-DNA-Moleküle Erfindung,
vorzugsweise in rekombinanter Form als Plasmide, direkt
oder als Bestandteil eines rekombinanten Virus, oder
Bakteriums verabreicht werden. Prinzipiell kann jede
gentherapeutische Methode für die Immuntherapie von
20 Krebs auf Basis von DNA ("DNA-Vakzine") auf B345-DNA
angewendet werden, und zwar sowohl *in vivo* als auch
ex vivo.

Beispiele für die *in vivo* Verabreichung sind die direkte
Injektion von "nackter" DNA, entweder intramuskulär oder
25 mittels Gen-Pistole ("gene gun") von der sich gezeigt
hat, daß sie zur Bildung von CTLs gegen Tumorantigene
führt. Beispiele für rekombinante Organismen sind
Vaccinia Virus, Adenovirus oder Listeria monocytogenes
(eine Übersicht wurde von Coulie, 1997, gegeben).
30 Des Weiteren können synthetische Träger für

Nukleinsäuren, wie kationische Lipide, Mikrosphären, Mikrokügelchen oder Liposomen für die *in vivo* Verabreichung von Nukleinsäure-Molekülen, kodierend für B345-Peptid verwendet werden. Ähnlich wie für Peptide können verschiedene Hilfsstoffe, die die Immunantwort verstärken, mitverabreicht werden, z.B. Zytokine, entweder in Form von Proteinen oder dafür kodierenden Plasmiden. Die Applikation kann gegebenenfalls mit physikalischen Methoden, z.B. Elektroporation, kombiniert werden.

Ein Beispiel für die *ex vivo* Verabreichung ist die Transfektion dendritischer Zellen, wie von Tuting, 1997, beschrieben, oder anderer APCs, die als zelluläre Krebsvakzine zur Anwendung kommen.

Die vorliegende Erfindung betrifft somit in einem weiteren Aspekt die Verwendung von Zellen, die B345 exprimieren, entweder von sich aus oder, in gegebenenfalls modifizierter Form, nach Transfektion mit der entsprechend kodierenden Sequenz, für die Herstellung einer Krebsvakzine.

Alternativ zur natürlichen B345-cDNA bzw. Fragmenten davon können modifizierte Derivate verwendet werden. Diese umfassen Sequenzen mit Modifikationen, die für ein Protein(fragment) bzw. Peptide mit stärkerer Immunogenität kodieren, wobei für die Modifikationen auf DNA-Ebene dieselben Überlegungen gelten wie für die oben beschriebenen Peptide. Eine weitere Art der Modifikation ist die Aneinanderreihung zahlreicher Sequenzen, kodierend für immunologisch relevante Peptide, nach Art einer Perlenschnur ("string-of-beads"; Toes et al.,

1997). Die Sequenzen können auch durch Anfügung von Hilfselementen modifiziert werden, z.B. Funktionen, die eine effizientere Abgabe und Prozessierung des Immunogens gewährleisten (Wu et al., 1995).

- 5 Beispielsweise kann durch Anfügen einer Lokalisierungssequenz in das endoplasmatische Retikulum ("ER targeting sequence") die Prozessierung und damit die Präsentation und letztlich die Immunogenität des Antigens erhöht werden.
- 10 Die vorliegende Erfindung betrifft in einem weiteren Aspekt ein rekombinantes DNA-Molekül, das B345-DNA enthält, z.B. verbunden mit einer regulatorischen DNA-Sequenz, insbesondere einer heterologen regulatorischen DNA-Sequenz, z.B. einem Promoter oder
- 15 Enhancer.

Die Erfindung betrifft in einem weiteren Aspekt Antikörper gegen B345 bzw. Fragmente davon. Polyklonale Antikörper können in herkömmlicher Weise durch Immunisierung von Tieren, insbesondere Kaninchen, mittels Injektion des Antigens bzw. Fragmenten davon, und anschließender Reinigung des Immunglobulins erhalten werden.

- 20 Monoklonale anti-B345-Antikörper können nach Standardprotokollen gemäß dem von Köhler und Milstein, 1975, beschriebenen Prinzip gewonnen werden, indem Tiere, insbesondere Mäuse, immunisiert, anschließend antikörperproduzierende Zellen der immunisierten Tiere immortalisiert werden, z.B. durch Fusion mit Myelomzellen, und der Überstand der erhaltenen Hybridome
- 25 mittels immunologischer Standard-Assays auf monoklonale

anti-B345-Antikörper gescreent wird. Für den therapeutischen oder diagnostischen Einsatz im Menschen können diese tierischen Antikörper gegebenenfalls auf herkömmliche Weise chimerisiert (Neuberger et al., 1984; 5 Boulian et al., 1984) oder humanisiert (Riechmann et al., 1988; Graziano et al., 1995) werden.

Humane monoklonale anti-B345-Antikörper (fragmente) können auch von sog. „Phage Display Libraries“ (Winter et al., 1994; Griffiths et al., 1994; Kruif et al., 10 1995; Mc Guiness et al., 1996) und mittels transgener Tiere (Brüggemann et al., 1996; Jakobovits et al., 1995) gewonnen werden.

Die erfindungsgemäßen anti-B345-Antikörper können in immunhistochemischen Analysen für diagnostische Zwecke 15 eingesetzt werden, oder als Therapeutikum in der Krebstherapie. (Ein Beispiel für die erfolgreiche Anwendung eines monoklonalen Antikörpers in der Krebstherapie ist Herceptin; ein Antikörper gegen das Proto-Onkogen HER2. Herceptin kann in Brustkrebs- 20 Patienten angewendet werden, die eine Überexpression von HER2 aufweisen.)

In einem weiteren Aspekt betrifft die Erfindung die Verwendung von B345-spezifischen Antikörpern, um beliebige Substanzen selektiv zu bzw. in einen Tumor zu 25 bringen, der B345 exprimiert. Beispiele für solche Substanzen sind zytotoxische Agenzien oder radioaktive Nuklide, deren Wirkung darin besteht, den Tumor Vorort zu schädigen. Aufgrund der relativ tumorspezifischen Expression von B345 sind dabei nur geringe 30 Nebenwirkungen zu erwarten. In einem weiteren Aspekt

können mit Hilfe von B345-Antikörpern Substanzen zur Sichtbarmachung von Tumoren, die B345 exprimieren, herangezogen werden. Dies ist für die Diagnose und die Bewertung des Therapieverlaufs von Nutzen.

5 Therapeutische und diagnostische Anwendungen von Antikörpern, die für anti-B345-Antikörper in Frage kommen, sind in der WO 95/33771 beschrieben.

Das Protein der Bezeichnung B345 gemäß der vorliegenden Erfindung und die davon abgeleiteten Proteinfragmente,

10 Peptide bzw. Peptid-Äquivalente oder Peptidomimetika können in der Krebstherapie eingesetzt werden, z. B. um eine Immunantwort gegen Tumorzellen zu induzieren, die die entsprechenden Antigen-Determinanten exprimieren. Vorzugsweise werden sie für die Therapie von
15 B345-positiven Tumoren verwendet, insbesondere beim Lungen- und Kolonkarzinom.

Es ist bekannt, dass tumorassoziierte Antigene tumorspezifische Mutationen aufweisen können, die zu einer immunologischen Unterscheidung zwischen Tumor und

20 Normalgewebe beitragen (Mandruzzato et al., 1997; Hogan et al., 1998; Gaudi et al., 1999; Wölfel et al., 1994). Um das Vorhandensein tumorspezifischer B345-Mutationen festzustellen, wird, zweckmäßig mit Hilfe von Sonden aus der, erfindungsgemäßen isolierten cDNA, die B345-cDNA aus
25 einem oder mehreren unterschiedlichen Tumoren kloniert und die erhaltenen Sequenzen mit Normalgewebs-B345-cDNA verglichen. Es werden Versuche durchgeführt, die zeigen sollen, ob Tumor-B345-Peptide aus einem gegenüber Normalgewebs-B345 mutierten Sequenzabschnitt im
30 Vergleich zu Normalgewebs-B345-Peptiden aus dem entsprechenden Abschnitt eine verstärkte Immunogenität

aufweisen. Um zu bestätigen, dass etwaige Mutationen tumorspezifisch sind, können Antikörper gegen diese Bereiche generiert und Tumorzellen auf Expression von möglichen Mutationen untersucht werden.

5 Die vorliegende Erfindung betrifft somit in einem weiteren Aspekt B345-Peptide, abgeleitet von Bereichen eines tumorexprimierten B345, die tumorspezifische Mutationen aufweisen.

Auf Grund der bevorzugten Expression von B345 in

10 Tumorzellen kann angenommen werden, dass dieses Protein eine wichtige Funktion für den Tumor hat, z.B. für Entstehung, Infiltration und Wachstum und somit ein Target für die chemotherapeutische Intervention darstellt.

15 Im Hinblick auf seinen Einsatz als Target in der gezielten Chemotherapie wird B345 näher charakterisiert, um die geeignete Strategie für die Intervention mit dieser Funktion zu entwickeln.

Als ersten Schritt bei der sog. „down-stream“

20 Funktionsanalyse von B345 führt man zweckmäßig in einem ersten Schritt eine bioinformatische Analyse durch, die den für die experimentelle Validierung von B345 als Target richtungweisend ist.

25 Für diese Analyse stellen die auf Ähnlichkeit und modularer Struktur beruhenden Bioinformatik-Konzepte eine wesentliche Grundlage dar. Etablierte bioinformatische Hilfsmittel zur Feststellung von Ähnlichkeiten sind BLAST

(<http://www.ncbi.nlm.nih.gov/BLAST>, Altschul et al.,

1997) oder FASTA (Pearson & Lipman, 1988), die spezialisierten Datenbanken wie Pfam (<http://www.sanger.ac.uk/Pfam>, Bateman et al., 2000) und SMART (<http://smart.embl-heidelberg.de>, Schultz et al., 2000), welche Domänenstrukturen berücksichtigen. Zur Verfeinerung der Analyse können Applikationen wie Clustal (<http://www2.ebi.ac.uk/clustalw>, Higgins et al., 1996) HMMer (<http://hmmer.wustl.edu>), PSI-BLAST (Altschul et al., 1997) und die PROSITE Datenbank (<http://www.expasy.ch/prosite>, Hofmann et al., 1999) herangezogen werden. Statistische Analysemethoden, die nicht auf Homologien beruhen, gestatten die Vorhersage weiterer struktur- und funktionsrelevanter Eigenschaften wie der Sekundärstruktur und des Auftretens von Transmembransegmenten und Helix-Turn-Helix-Motiven. Methoden zur Vorhersage der Sekundärstruktur von Proteinen sind verfügbar; besonders erwähnenswert ist Jpred (<http://barton.ebi.ac.uk/servers/jpred.html>, Cuff et al., 1998). Die Sekundärstrukturvorhersage kann Funktionshypthesen untermauern, etwa wenn die Struktur des vermuteten Homologen bekannt ist.

Gemäß Bioinformatikanalyse weist B345 eine helikale Transmembrandomäne auf, wobei sowohl der N-terminale als auch der C-terminale Bereich hydrophile sind, was darauf schließen lässt, dass dieses Protein ein Transmembranprotein darstellt. Der N-terminale, extrazelluläre Bereich besitzt einige CUB-Domänen, welche zu Disulfidbrückenbildung neigen und daher bei der Dimerisierung bzw. bei Protein-Protein Wechselwirkungen beteiligt sind (Bork et al., 1993). Das C-terminale, intrazelluläre Ende zeigt Homologien zu einer Rezeptorkinase und zu einem C-Kinase Substrat.

In weiterer Folge wird B345 einer biochemischen und biologischen Analyse unterworfen.

In einem nächsten Schritt wird die Funktion von B345 für das Tumorgeschehen aufgeklärt; z.B. durch

- 5 Proliferationsassays in vitro oder in Tiermodellen, die das zu untersuchende B345-Gen überexprimieren (konstitutiv oder induzierbar) und als Kontrolle entweder in deletierter (inaktiver) Form exprimieren oder über Antisense hinunterregulieren (siehe z.B.
- 10 Grosveld und Kollias, 1992).

B345 kann in Screening-Assays verwendet werden, um Substanzen zu identifizieren, die die Aktivität dieses Proteins modulieren, insbesondere inhibieren. In einer Ausführungsform kann ein derartiger Assay z. B. darin

- 15 bestehen, das B345 Protein, oder ein aktives Fragment davon, in Zellen, die auf die Aktivität von B345 mit Proliferation reagieren, einzubringen bzw. die entsprechende B345 cDNA in der Zelle zur Expression zu bringen, und die Proliferation der Zellen in Gegenwart
- 20 und in Abwesenheit einer Testsubstanz zu bestimmen.

Ein Beispiel für Testzellen sind Zellen mit niedriger Teilungsrate, z.B. primäre Zellen, die kein endogenes B345 aufweisen. Um die Eignung der Zellen für einen Screening-Assay festzustellen, werden diese mit

- 25 B345-cDNA transformiert, gezüchtet und mit Standard-Assays, z.B. Thymidin-Einbau, auf ihre Proliferationsfähigkeit getestet. Aufgrund einer nach B345-Expression signifikanten Erhöhung ihrer Proliferationseigenschaft können sie als Testzellen eingesetzt werden, z.B. in High Throughput Screening
- 30

Proliferationsassays. Beispiele für Proliferationsassays im High Throughput Format, z.B. auf Grundlage des MTS-Assays, sind in der WO 98/00713 beschrieben.

Substanzen mit proliferationshemmender Wirkung können
5 zur Behandlung von Tumoren mit starker B345-Expression verwendet werden, insbesondere beim Lungen- und Kolonkarzinom.

Figurenübersicht:

10 Fig. 1A: Expressionsprofil von B345, B452 und B540 in individuellen Lungenkarzinomen und Lungentumorzelllinien.

Fig. 1B: Expressionsprofil von B345 in normalem Dickdarmgewebe und Tumorzelllinien.

15 Fig. 1C: Graphische Darstellung des Alignments von B345, B452 und B540.

Fig. 2A: Northern Blot Analyse der Tumorzelllinie A549 mit einem 490bp langem B345 PCR-Produkt

20 Fig. 2B: Northern Blot Analyse verschiedener Normalgewebe mit einem 490bp langem B345 PCR-Produkt

Fig. 2C: Northern Blot Analyse verschiedener Krebsgewebe mit einem 318bp langen B345 PCR-Produkt

25 Fig. 3 mRNA Expressionsanalyse von B345 durch real-time PCR von Tumor- und Normal-Geweben.

Fig. 4: mRNA Expressionsanalyse von B345 durch real-time PCR von Laser-Mikroskop-präparierten Dickdarmtumoren (LCM) sowie normalem Dickdarmgewebe und Tumorzelllinien.

5 Fig. 5: Graphische Darstellung der Genstruktur von B345.

Fig. 6: Hydrophilizitäts-und Transmembran-Blot des B345-Proteins

Fig. 7: Potentielle Proteinstruktur von B345

10

Tabellenübersicht

Tab. 1: Zusammenfassung der Northern Blot Daten von B345 in verschiedenen Normalgeweben (1A),

15 Krebszelllinien (1B); und verschiedenen Normalgeweben im Vergleich mit dem entsprechenden Tumorgewebe (1C)

20 Tab. 2A: Zusammenfassung der Daten der quantitativen PCR von B345 in verschiedenen Normal- und Krebsgeweben

Tab. 2B: Zusammenfassung der Daten der quantitativen PCR von B345 in verschiedenen Normalgeweben und mikrodissezierten Kolonadenokarzinom Geweben

25

Zeichenerklärung

+++	extrem positiv
++	stark positiv
+	positiv
5 (+)	schwach positiv
-	negativ

Beispiel 1

10 RDA („Representational Difference Analysis“) von der humanen Adenokarzinom-Zelllinie der Lunge (A549) und normalem Lungengewebe

Die von ATCC bezogene humane Lungenadenokarzinom-Zelllinie A549 (CCL 185) wurde in T150 Zellkulturflaschen hochgezüchtet. Als Nährmedium diente MEM mit 10% hitze-15 inaktiviertem, fötalem Kälberserum und 2 mM L-Glutamin. Alle 3 bis 4 Tage wurden die Zellen durch Trypsinisieren 1:5 bis 1:10 zur Propagation gespalten. Nach Erreichen von etwa 80% Konfluenz wurden pro T150 Zellkulturflasche 4 ml einer Trypsinlösung (Angaben pro Liter: 8g NaCl, 20 0,2g KCl, 1,13g Na₂HPO₄-wasserfrei, 0,2g KH₂PO₄, 100ml 2,5% Trypsinlösung, 1g EDTA-Na-Salz; pH 7,2 - 7,4) zum Ernten der Zellen eingesetzt. Die 4 ml wurden in ein 15 ml Falconröhrchen transferiert, mit 8 ml PBS versetzt, bei 1200 rpm in einer Haereus Tischzentrifuge 25 (Megafuge 2.0R) 5 min bei 4°C zentrifugiert, das Zellpellet mit 1 ml Lysis-Puffer (10mM Tris-HCl pH8, 140mM NaCl, 1,5 mM MgCl₂, 0,5% NP40) versetzt, kräftig

geschüttelt und in einem 2 ml Eppendorfgefäß bei
12 000 rpm und 4°C 5 min in einer Sigma Tischzentrifuge
(Sigma 202 MK) abzentrifugiert. Der Überstand wurde in
ein neues Eppendorfgefäß transferiert und nach Zusatz
5 von 55 µl 20% SDS-Lösung zweimal mit dem doppelten
Volumen an einer CHCl₃/Phenol (1:1 v/v) Mischung und
einmal mit dem einfachen Volumen an CHCl₃ extrahiert. Die
wässrige RNA-enthaltende Phase wurde mit 1/10 Volumen an
3M NaAc (pH5) und dem zweifachen Volumen an 96% EtOH
10 versetzt und über Nacht bei -20°C die RNA präzipitiert.
Ausgehend von 1 mg Gesamt-RNA wurde für die Isolierung
von poly-A(+)RNA mittels des PolyATtract Kit (Promega)
entsprechend dem Hersteller-Protokoll vorgegangen. Die
Lagerung der A549 poly-A(+)RNA mit einer Konzentration
15 von 1 mg/ml in DEPC-behandeltem H₂O erfolgte in Aliquots
bei -80°C.

Zur Durchführung der repräsentativen Differenzanalyse
(RDA; Hubank and Schatz, 1994; Diatchenko et al., 1996)
wurde die poly-A(+)RNA der Lungenadenokarzinom-Zelllinie
20 A549 als „tester“, die von normalem Lungengewebe
(1 mg/ml; Clontech, Palo Alto; #6524-1) als „driver“
eingesetzt. Die RDA wurde unter Verwendung des
PCR-select™ kit (Clontech, Palo Alto) entsprechend dem
Hersteller-Protokoll durchgeführt, mit der Ausnahme,
25 dass ein modifiziertes Primer/Adaptor-2-
Oligonukleotidsystem zum Einsatz kam: Adaptor-2-alt-1
(SEQ ID NO:31) und nested-PCR-primer-2-alt
(SEQ ID NO: 32) und Adaptor-2-alt-2 (SEQ ID NO: 33). Die
neu generierten Primer/Adaptor-Sequenzen ermöglichen
30 durch die Anwesenheit von drei neuen
Restriktionsenzymschnittstellen (Kpn I, Sac I und Xho I)
in der Sequenz des nested-PCR-primer-2-alt nach

Klonierung der subtrahierten cDNA-Fragmente in den pCRII-Vektor ein nachträgliches Herausschneiden der jeweiligen cDNA-Fragmente. Das Designen einer Primer/Adaptorsequenz mit mehreren verfügbaren

5 Restriktionsenzymschnittstellen war deshalb notwendig, weil besonders in den Primersequenzen - bedingt durch die PCR-Amplifikationsschritte- oft Punktmutationen zu beobachten waren.

Nach der Synthese von doppelsträngiger cDNA mittels

10 oligo-dT wurde die erhaltene cDNA von „tester“ und „driver“ mit *Rsa*I verdaut (*Rsa*I ist ein 4-Basen erkennendes Restriktionsenzym und liefert im statistischen Mittel 256 bp lange cDNA-Fragmente). Gleiche Teile von „tester-cDNA“ wurden entweder mit den 15 Adaptoren 1 oder 2 ligiert und anschließend getrennt mit einem Überschuss an „driver-cDNA“ bei 65°C hybridisiert. Danach wurden die beiden Ansätze vereinigt und einer zweiten Hybridisierung mit frischer denaturierter „driver-cDNA“ unterworfen. Die angereicherten „tester“- 20 spezifischen cDNAs wurden anschließend durch PCR, mit für die Adaptoren 1 bzw. 2 spezifischen Primern, exponentiell amplifiziert. Für eine weitere Anreicherung wurde ein Aliquot dieser Reaktion einer zweiten PCR mit spezifischen nach innen versetzten („nested“) Primern 25 unterworfen. Die aus dieser Reaktion resultierenden, exponentiell amplifizierten cDNA-Fragmente wurden direkt in den pCRII-Vektor (Invitrogen; „TA-cloning vector“) ligiert und anschließend ein Drittel des Ligationsansatzes in kompetente *E. coli* (OneShot™, 30 Invitrogen) transfiziert.

712 positive Transformanten (Blau-Weiß-Selektion) wurden erhalten und in 96-Napf Blöcken in LB-Amp Medium (1,3 ml pro Napf) für 48 h bei 37°C kultiviert. Pro Napf wurden 750 µl der *E. coli* Suspensionen für die Präparation der 5 Plasmid-DNA eingesetzt (96-Napf- Minipräparationsmethode von QIAgen nach Vorschrift des Herstellers). Die verbleibenden Bakterienkulturen wurden als Glycerinstammkulturen bei -80°C gelagert.

Es wurde eine aus 712 Einzelklonen bestehende cDNA-
10 Subtraktionsbibliothek erhalten, die sowohl in Form von *E. coli* Glycerin-Stammkulturen als auch in Form gereinigter Plasmide vorlag.

Beispiel 2

15 DNA-Sequenzierung und Annotation von TAA-Kandidaten

Die isolierte Plasmid-DNA sämtlicher 712 Klone (siehe Beispiel 1) wurde nach der Sanger-Methode auf einem ABI-377 Prism Gerät sequenziert. Die erhaltenen Sequenzen wurden mittels der BioScout-Software (LION, Heidelberg) 20 annotiert und Datenbankvergleichen (Genbank) unterzogen. Von 712 Klonen konnten 678 sequenziert und annotiert werden. Der Rest (34) wies als Insert entweder nur poly(A) Sequenzen auf oder entsprach einem religierten Vektor oder war nicht sequenzierbar. Von den 25 678 annotierbaren Sequenzen erwiesen sich 357 als Gene mit bekannter Funktion. Die restlichen 321 repräsentierten Klone, kodierend für Gene mit unbekannter Funktion; 59 davon wiesen nicht einmal

Einträge in der humanen EST-Datenbank auf. Bekannte Gene wurden nicht weiter behandelt. Für jene unbekannten Gene, für die ein EST-Eintrag verfügbar war, wurde eine Abschätzung des Expressionsprofils vorgenommen: dabei

5 wurden all jene ESTs mit > 95% Identität (BLAST), die zur entsprechend experimentell ermittelten Sequenz der Subtraktionsbibliotheken gehörten, überprüft. Bei der Annotation wurde eine Unterteilung in a) kritische Normalgewebe, b) fötale, "verzichtbare" und

10 immunprivilegierte Gewebe und c) Tumore und Tumorzelllinien vorgenommen. Auf der Basis dieses "virtuellen mRNA-Profiles" ("virtueller Northern blot") wurden 200 Klone, für die keine ESTs in der Gruppe a) gefunden wurden, für weitere experimentelle Analysen ausgewählt

15 (inklusive der 59 Klone, für die keine EST-Eintragung vorlag). Zur weiteren Einengung der Kandidatenklone wurden aus den von den 200 ausgewählten Klonen ermittelten Sequenzen Oligonukleotidprimerpaare entworfen und synthetisiert. Es wurden zunächst

20 8 verschiedene, von humanem Gewebe abgeleitete cDNA-Bibliotheken (GibcoBRL "SUPERSCRIPT™"), welche direktional in pCMV-SPORT kloniert sind, mittels qualitativer PCR auf das Vorhandensein der jeweiligen Kandidaten getestet. Die dabei eingesetzten

25 cDNA-Bibliotheken stammten aus Gewebe von Herz (#10419-018), Leber (#10422-012), Leukozyten (#10421-022), Niere (#10420-016), Lunge (#10424-018), Testis (#10426-013), Gehirn (#10418-010) und fötalem Gehirn (#10662-013). Die PCR-Bedingungen waren wie

30 folgt: 20 µl Gesamtvolumen pro PCR-Ansatz enthielten

1 x TaqPol-Puffer (50mM KCl, 10 mM Tris-HCl pH9,
0,1% Triton X-100), 1,5 mM MgCl₂, 0,2 mM dNTPs (Promega),
0,025 U/ μ l Taq-DNA-Polymerase (Promega), jeweils 5 pM an
spezifischen Oligonukleotidprimer für B345 (B345-D,

5 SEQ ID NO: 34) und (B345-U, SEQ ID NO: 35) sowie 100 ng
der jeweils zu untersuchenden Plasmid-DNA. Zur Kontrolle
wurden spezifische Primer für GAPDH (SEQ ID NO: 36 und
37) eingesetzt. Zur Überprüfung des selektiven
Nachweises wurden die jeweiligen B345 spezifischen

10 Primerpaare Oligonukleotidprimer (SEQ ID NO: 34) und
(SEQ ID NO: 35) parallel auch auf das isolierte Plasmid
mit dem B345 "original Fragment" hin ausgetestet
(ursprünglich isoliertes cDNA-Fragment von B345). Die

Nachweisbarkeit von Fragmenten der zu erwartenden Länge
15 mit starkem Signal in einem der kritischen Normalgewebe
(Herz, Leber, Lunge, Niere und Leukozyten), nicht jedoch
in den cDNA-Bibliotheken aus immunprivilegierten Geweben
(Gehirn, fötales Gehirn und Testis) unter diesen

PCR-Bedingungen (1 Zyklus: 3' 94°C; 35 Zyklen: 1' 94°C -
20 1' 55°C - 1' 72°C; 1 Zyklus: 7' 72°C) wurde als
Ausscheidungskriterium definiert. Mittels dieser
qualitativen PCR-Analyse konnte die Zahl der Kandidaten
auf 56 eingeengt werden; Klon B345 befand sich in dieser
bereits vorselektionierten Kandidatengruppe.

25

Beispiel 3

Expressionsanalyse durch cDNA Chip Hybridisierung

Für das Design eines cDNA Chips wurden von der dBEST Datenbank über Nukleotid-Sequenzsuche eine Reihe von 5 Klonen ausgewählt, die zu verschiedenen Funktionskategorien, wie Apoptose bis zur Zellzyklus-Regulation, gehören. Insgesamt wurden 1299 IMAGE Klonen (davon repräsentieren 1024 bereits bekannte Gene) bestellt und zur Kontrolle sequenziert.

10 Mikrotiterplatten mit Bakterien, die ca. 800 bp lange Sequenzen vom 3`Ende des Gens im Vektor enthalten, wurden an Incyte Pharmaceuticals, Inc. (USA) geschickt und diese dort auf 60 Chips gespottet. Neben diesen Klonen wurden auch 120 durch RDA identifizierten EST 15 Klone auf die Chips gespottet. Die so produzierten DNA Chips wurden anschließend mit Cy3 markierter cDNA aus Normalgewebe, Tumorgewebe und Zelllinien zusammen mit Cy5 markierter cDNA aus einer Mischung von neun verschiedenen Normalgeweben hybridisiert und die beiden 20 Signale zur Normalisierung der Expressionswerte verglichen. Die Berechnungen erfolgten teilweise in S-Plus oder in Microsoft Excel. Die Auswertung der Chip Experimente ergab ein sehr ähnliches Expressionsprofil für B345, B540 und B452 bei der Hybridisierung mit 25 Lungenkrebs Proben von Zelllinien und Patientenmaterial (Siehe Fig. 1A). Solch ein tumorassoziiertes Expressionsprofil konnte für B345 auch bei Vergleich von Kolon Adenokarzinom mit Kolon Normalgewebe gezeigt werden (Siehe Fig. 1B).

Das Sequenzalignment von B345, B540 und B452 zeigte eindeutig ein Überlappen der einzelnen EST Fragmente. Daher konnte man annehmen, dass es sich bei den drei Klonen um ESTs von ein und dem selben Gen handelt. Der daraus resultierende DNA Abschnitt deckt eine Länge von 843 bp ab (Siehe Fig. 1C) und wurde in weiteren Versuchen zur Durchsuchung von öffentlichen Datenbanken verwendet. Die Suchergebnisse lieferten keine signifikante Homologie zu bekannten DNA bzw. Protein Sequenzen, was darauf hindeutet, dass es sich bei B345 um ein bis lang unbekanntes Gen handelt.

Beispiel 4

Expressionsanalyse von B345 mittels Northern Blots

Bei B345 handelt es sich um ein Gen, das laut DNA Chip Analysen in Tumorgeweben (siehe Fig. 1A und 1B, Tab. 1a und Tab. 1B) hochreguliert ist.

Um einerseits das erhaltene Transkriptionsprofil zu bestätigen und andererseits die Länge der zu erwartenden mRNA für die full size Klonierung zu bestimmen, wurde für B345 eine Northern Blot Analyse unter Einsatz von humanen Zelllinien und des "Human Multiple Tissue Northern Blots" (Clontech und Invitrogen) durchgeführt. Als Sonden dienten die mit [α -³²P]dCTP (NEN, Boston) markierten 490 bp bzw. 318bp langen PCR Produkte von B345 (Primer (SEQ ID NO: 5 und SEQ ID NO: 6 bzw. SEQ ID NO:7 und SEQ ID NO:8)). Die Hybridisierung erfolgte bei 68° für 2 h; die Visualisierung durch Standard-

Autoradiografie (Hyperfilm, Amersham). Fig. 2A, 2B und 5 Tab. 1a und Tab. 1B, 1C zeigen das Ergebnis dieser Analyse: Fig. 2A von der Zelllinie A549, Fig. 2B von 12 Normalgeweben (Peripherie Blut Lymphozyten (PBL), Lunge, Placenta, Dünndarm, Leber, Niere, Milz, Thymus, Kolon, Skelettmuskel, Herz und Hirn) und Fig. 2C von 10 8 Krebszelllinien (Promyelozytische Leukämie HL60, HeLa-S3, chronische myelogene Leukämie K-562, lymphoblastische Leukämie MOLT-4, Burkitt's Lymphom (Raji), Kolon Adenokarzinom SW480, Lungen Adenokarzinom A549 und Melanom G361). Das B345-Transkript zeigt eine Länge von 6,5 kb.

Beispiel 5

15 Analyse des Expressionsprofils von B345 auf RNA-Ebene mittels quantitativer RT-PCR (real time PCR oder TaqMan-Analyse)

Um eine exaktere Quantifizierung der mRNA Expression in den verschiedenen normal und Tumorgeweben durchzuführen, 20 wurde die "real time PCR" angewandt, die es erlaubt die RNA Konzentration im Vergleich zu einen externen Standard zu berechnen.

Die Isolierung der RNA aus Gefriergewebe erfolgte mit Trizol gemäß dem Herstellerprotokoll von Gibco. Zur Entfernung von etwaig kontaminierender DNA wurde die präparierte RNA wie folgt mit DNAase I verdaut: 3 µg Gesamt-RNA wurden mit 20 µl 5x AMV Puffer (Promega), 25 1 µl RNasin (Promega) und 2 µl DNase I (Boehringer

Mannheim) in einem Gesamtvolumen von 80 µl 15 Minuten bei 37°C inkubiert. 120 µl Phenol : Chloroform : Isoamylalkohol (25:24:1) wurden zugegeben, auf einem Vortexer gemischt und kurz abzentrifugiert. Die wässrige

5 Phase wurde abgenommen, mit 120 µl Chloroform: Isoamylalkohol (24:1) versetzt und wie vorher zentrifugiert. Die gereinigte RNA wurde Ethanol-gefällt und in Wasser gelöst.

Anschließend wurde die Gesamt-RNA mit reverser

10 Transkriptase (Superscript, Gibco, BRL) in cDNA umgeschrieben: Zu 3 µg Gesamt-RNA wurden 1 µl Oligo dT primer (Promega) zugegeben und mit Wasser auf ein Endvolumen von 10 µl gebracht. Nach einer Inkubation von 5 Minuten bei 70°C wurde die Lösung 5 Minuten bei
15 Zimmertemperatur abgekühlt. 5 µl RT reaction buffer (5x, Gibco, BRL), 2,5 µl dNTPs (je 10 mM, Boehringer Mannheim), 1 µl RNasin (10U/µl, Promega), 1,5 µl Superscript (10 U/µl, Gibco, BRL) und 5 µl Wasser wurden zugegeben und 1 Stunde bei 42°C inkubiert und die
20 Reaktion durch Inkubation von 3 Minuten bei 95°C beendet.

Zur Herstellung eines cDNA-pools eines bestimmten Gewebe oder Tumortyps wurden 3 bis 10 verschiedene Einzelpräparationen von unterschiedlichen Patienten in

25 gleichen Anteilen gemischt.

Die quantitative Bestimmung der "Haushaltsgene" β-Aktin, GAPDH und Tubulin in cDNA-pools wurde wie folgt durchgeführt:

A) β -Actin-TaqMan PCR (Perkin Elmer)

Details über das Prinzip der TaqMan Methode siehe Herstellerinformation (Perkin Elmer). Ein TaqMan PCR Lauf beinhaltete Proben an β -Actin-Kontrollsequenz mit 5 je 10^2 , 10^3 , 10^4 , 10^5 und 10^6 Kopien/ μ l (Perkin Elmer) zur Bestimmung der Standardkurve, eine Negativkontrolle ohne DNA und die zu quantifizierenden cDNA-pools. Alle Proben wurden in Triplikaten analysiert. Für einen 25 μ l Reaktionsansatz wurden 1 μ l cDNA, 2,5 μ l 10x Puffer A 10 (Perkin Elmer), 4 μ l MgCl₂ (25 mM, (Perkin Elmer)), 0,5 μ l je Nukleotid (10 mM dATP, dCTP, dGTP; 20 mM dUTP), 0,125 μ l TaqMan Sonde (20 μ M; TaqMan Sonde für β -Aktin (SEQ ID NO: 20 fluoreszenzmarkiert am 5'-Ende mit 15 6-Carboxyfluorescein und mit 6-Carboxytetramethylrhodamine am 3'-Ende), 1 μ l je β -Aktin spezifischer Primer (je 20 μ M, Forward Primer SEQ ID NO:21 und Reverse Primer SEQ ID NO:22), 0,25 μ l AmpErase uracil N-Glycosylase "UNG" (1 U/ μ l, Perkin 20 Elmer), und 0,125 μ l AmpliTaq Gold (5 U/ μ l, Perkin Elmer) gemischt, in MicroAmp Optical Tubes (Perkin Elmer) überführt und mit MicroAmp Optical Caps verschlossen. Die PCR wurde folgendermaßen durchgeführt: ein Zyklus mit 2 Minuten 50°C für die UNG Reaktion, ein Zyklus 10 Minuten 95°C zur Aktivierung der AmpliTaq, 25 40 Zyklen mit je 15 Sekunden 95° und 1 Minute 60°C. Anschließend wurden die Proben auf 25°C gehalten. Die Auswertung der Daten erfolgt mit dem Programm "Sequence Detection System 1.5b1" (PE Applied Biosystems), wobei im Prinzip die Fluoreszenzsignale der zu 30 quantifizierenden cDNA-Proben mit den Signalen der

Kontrollplasmidverdünnungen bekannter Konzentration verglichen wurden.

B) GAPDH-TaqMan PCR

Für die Quantifizierung von GAPDH, das wie β -Aktin oder Tubulin zur Normalisierung der eingesetzten RNAs verwendet wurde, sind folgende Primer bzw. Sonden eingesetzt worden. Als TaqMan Sonde für GAPDH diente (SEQ ID NO: 23) eine am 5'-Ende mit Tetrachlorfluorescein und am 3'-Ende mit Carboxymethylrhodamin markierte Sonde (Forward GAPDH Primer: SEQ ID NO: 24 und Reverse Primer: SEQ ID NO: 25). Die Reaktionen wurden wie oben beschrieben durchgeführt.

C) Tubulin-SybrGreen PCR (Perkin Elmer)

Prinzip der SybrGreen PCR siehe Herstellerinformation (Perkin Elmer). Ein SybrGreen PCR Lauf beinhaltete Proben an Tubulin-Kontrollplasmid mit je 10^2 , 10^3 , 10^4 , 10^5 und 10^6 Kopien/ μ l (Perkin Elmer) zur Bestimmung der Standardkurve, eine Negativkontrolle ohne DNA und die zu quantifizierenden cDNA-pools. Alle Proben wurden in Triplikaten analysiert. Für einen 25 μ l Reaktionsansatz wurden 1 μ l cDNA, 2,5 μ l 10 \times SybrGreen Puffer (Perkin Elmer), 3,5 μ l MgCl₂ (25 mM, Perkin Elmer), 0,5 μ l je Primer (je 20 μ M, Perkin Elmer, Tubulin Forward (SEQ ID NO:26); Tubulin reverse (SEQ ID NO:27), 0,25 μ l AmpErase uracil N-Glycosylase "UNG" (1 U/ μ l, Perkin Elmer), und 0,25 μ l AmpliTaq Gold (5 U/ μ l, Perkin Elmer) gemischt, in MicroAmp Optical Tubes (Perkin Elmer) überführt und

mit MicroAmp Optical Caps verschlossen. Die PCR wurde folgendermaßen durchgeführt: ein Zyklus mit 2 Minuten 50°C für die UNG Reaktion, ein Zyklus 10 Minuten 95°C zur Aktivierung der AmpliTaq, 40 Zyklen mit je 15
5 Sekunden 95° und 1 Minute 60°C. Anschließend wurden die Proben auf 25°C gehalten. Die Auswertung der Daten erfolgt mit dem Programm "Sequence Detection System 1.5b1" (PE Applied Biosystems), wobei im Prinzip die Fluoreszenzsignale der zu quantifizierenden cDNA-Proben
10 mit den Signalen der Kontrollplasmidverdünnungen bekannter Konzentration verglichen wurden.

D) B345-TaqMan PCR

Die quantitative TaqMan-PCR-Analyse von B345 wurde, wie für die "Haushaltsgene" beschrieben, durchgeführt. Es wurden jedoch B345 spezifische Primer (SEQ ID NO:28 und SEQ ID NO:29) (200 ng/µl) und eine B345 spezifische Sonde (SEQ ID NO:30, 20 µM), die am 5'-Ende mit Tetrachlorfluorescein und am 3'-Ende mit Carboxymethylrhodamin markiert ist, verwendet. Als Standard wurde das PCR Produkt von B345 mit den Primern SEQ ID NO: 28 und SEQ ID NO: 29 mit bekannter Kopienzahl eingesetzt.
20

Fig. 3 veranschaulicht die TaqMan- Expressionsanalyse (Fig. 3A: β-Actin; Fig. 3B: Tubulin). Es zeigte sich, dass B345 höher in Dickdarmkrebsgewebe als in Normalgewebe exprimiert ist (siehe Tab. 2a). Nun stellt aber sowohl das Normalgewebe als auch das Tumorgewebe ein sehr heterogenes Gemisch von verschiedenen Zelltypen dar. Weiteres variiert der Anteil von Tumorzellen im
25

Tumorgewebe sehr stark von etwa 30 bis 80%. Um diese biologische Heterogenität auf ein Minimum zu beschränken, wurden die Epithelzellen des Dickdarms, die die Ursprungszellen des Adenokarzinoms darstellen, und 5 die Krebszellen oder Krebsareale durch Laser-Mikrodissektion spezifisch präpariert. Gewebeschnitte von 10µm Stärke wurden mit dem Kyromikrotom von Leica, Jung CM1800 angefertigt und auf einen mit Polyethylen beschichteten Objektträger aufgebracht (Böhm et al., 10 1997). Die bei Raumtemperatur für etwa 30 Minuten getrockneten Schnitte wurden mit Mayers Hämatoxylin (SIGMA DIAGNOSTICS) inkubiert und anschließend zur Entfernung von unspezifisch gebundenen Farbstoff fünf Minuten unter fließendem Wasser gewaschen. Nach fünf 15 minütigem Trocknen bei 37°C wurde die Laser-Mikrodissektion durchgeführt. Dafür wurde das Lasermikroskop von PALM (PALM GmbH, Bernried, Deutschland) eingesetzt und etwa 2000 bis 5000 Zellen präpariert. Die durch Reverse Transkription gewonnene 20 cDNA, wurde auch hier durch Real Time PCR analysiert. Das Ergebnis zeigt, dass die B345 Expression in Dickdarmkarzinom Zelllinien sowie in Patientenmaterial um ein Vielfaches höher ist als vergleichsweise die des Dickdarmnormalgewebes. Zur Normalisierung wurde der 25 Expressionslevel von GAPDH bestimmt (siehe Fig. 4 und Tab. 2B).

Beispiel 6**a) Klonierung der cDNA von B345**

Das Durchsuchen von Datenbanken nach Sequenzen von Genfragmenten (ESTs, expressed sequence tags), die für die „in silico“ Klonierung von B345 herangezogen werden können, ergab ein überlappendes EST-Kontig von etwa 1500 bp. Der polyA-Bereich an einem der Enden gab Hinweis auf die Orientierung des DNA Abschnittes in Bezug auf 5' - 3' Orientierung, was beim Design von neuen Primern für die Amplifikation von B345-spezifischen cDNA-Fragmenten essentiell ist.

Zunächst wurde das durch die Datenbankanalyse beschriebene potentielle 3' Ende durch experimentelle Ansätze verifiziert. RNA von der Lungen-Karzinom Zelllinie Calu 6 (AACC No. HTB56) wurde mit Hilfe des Primers (SEQ ID NO:9) revers transkribiert und die resultierende einzelsträngige cDNA wurde mittels PCR mit den Gen spezifischen Primer SEQ ID NO:5 und den Adapterprimer SEQ ID NO: 10 amplifiziert.

Für einen 25 µl PCR Ansatz wurden 1 µl des cDNA-pools mit 2,5 µl 10×Taq Puffer (Promega), 1,5 µl MgCl₂ (25 mM, Promega), 0,5 µl dNTPs (je 10 mM, Boehringer Mannheim), 1 µl Primermischung (je 20 µM), 0,15 µl Taq Polymerase (Promega) in Wasser gemischt. Die PCR wurde wie folgt durchgeführt: 1× 94°C 3 Minuten; 30× 94°C 30 Sekunden, 55°C 30 Sekunden, 72°C 1 Minute; auf 4°C halten. Die PCR wurde auf einem 1,2% Agarosegel analysiert.

Die beiden Primer wurden anschließend zum Sequenzieren des gereinigten PCR Produktes verwendet. Die ermittelten Sequenzen zeigten eine hohe Homologie mit dem „*in silico*“ klonierten DNA Abschnitt (inklusive des PolyA-Trakts).

Da das Klonieren von 5'-Endsequenzen einen meist sehr aufwendigen Prozess darstellt, wurden im Folgenden zur Lösung des Problems verschiedene Methoden angewandt.

Als Ausgangszelllinie wurde auch hier Calu 6 verwendet.

10 Nach der reversen Transkription der RNA mit dem Primer SEQ ID NO:9 und der Zweitstrangsynthese wurde an die Doppelstrang-cDNA ein Linker bestehend aus den beiden Oligos SEQ ID NO:11 und SEQ ID NO:12 ligiert (Abe et al., 1992). Die daraus resultierende LoneLinker cDNA
15 Bibliothek wurde dann mit dem Gen spezifischen Primer SEQ ID NO:6 linear über 35 Zyklen amplifiziert. Ein Aliquot der B345 angereicherten cDNA konnte anschließend mit den Primern SEQ ID NO: 13 und LLEcoRIA SEQ ID NO:11 weiter amplifiziert werden. Nach der Gelelektrophorese
20 eines Aliquots und Southernanalyse mit dem genspezifischen Oligo SEQ ID NO: 14 konnte eine 5 kb Bande lokalisiert werden. In weiterer Folge wurde dieses Fragment schrittweise sequenziert und auf die Sequenz des EST-Kontigs ausgerichtet ("aligned").

25 Um die resultierende Sequenz aus der LLcDNA Klonierung zu überprüfen, wurden zwei Fragmente mittels PCR (SEQ ID NO:15 und SEQ ID NO:16 bzw. SEQ ID NO:6 und SEQ ID NO:17) amplifiziert und für das Screenen von Lambda gt10 cDNA Phagen Bibliotheken eingesetzt. Positive Plaques
30 wurden isoliert und mittels den gt10 spezifischen

Primern (SEQ ID NO:18 und SEQ ID NO:19) PCR amplifiziert. Anschließende Sequenzierung und das Alignment mit den Sequenzen führte zu der Annahme, das es sich hierbei um ein differenzielles Spleißprodukt handelt. Die Spleißdonor, Akzeptor und Lariatsequenz konnten in weiterer Folge gefunden werden. Mittels PCR durch geeigneter Primerkombination wurde in verschiedenen Zelllinien nach differenziellen Spleißprodukten gesucht, wobei in allen gescreenten Zelllinien nur ein Produkt gefunden wurde und zu der in Fig. 5 dargestellten Genstruktur führte. Die angeführte cDNA besitzt einen offenen Leserahmen (ORF) der für ein potentielles Protein mit einer Länge von 749 Aminosäuren kodiert. Die Translationsinitiationsstelle an Position 215 entspricht zu etwa 75% einer Kozak-Konsensussequenz. Die in diesem Experiment erhaltenen Ergebnisse veranlassten dazu, in einem weiteren Experiment (Beispiel 6b) die Transkriptionsinitiationsstelle noch exakt durch Primer-Extension zu bestimmen, um sicher zu gehen, dass das hier ermittelte 5' Ende das tatsächliche 5'-Ende von B345 ist. Die von der in diesem Klonierungsversuch erhaltenen cDNA (SEQ ID NO:1) abgeleitete Aminosäuresequenz von B345 ist in SEQ ID NO:2 gezeigt.

b) Zweiter Klonierungsversuch; Bestimmung der 5'-und der Promotorregion von B345

Mit Hilfe des Promotor Finder DNA Walking Kit (Clontech) und anschließender Primerextension Reaktion wurden die 5'-Region und die Promotorregion sowie die exakte Transkriptionsinitiationsstelle bestimmt. Die 5'-Region

wurde mit Hilfe einer genomischen DNA Library von Clontech mit B345 spezifischen Primern (SEQ ID NO:38 bzw. nested SEQ ID NO:39) und Adaptor Primer im Kit amplifiziert. Um den exakten Transkriptionsstart zu bestimmen, wurde die Primer Extension Reaktion durchgeführt. Dafür wurde der Primer SEQ ID NO:40 am 5`-Ende mit Hilfe 10 U der T4 Polynukleotid Kinase (Promega) und 3 μ l [γ -³²P]ATP (3000Ci/mmol) nach Standard-Protokollen markiert (Sambrook et al., 1989). Das markierte Oligonukleotid wurde durch Präzipitation gereinigt. Für die Primerextension Reaktion wurden 10.000 cpm Oligonukleotid in einem Gesamtvolumen von 20 μ l zu 25 μ g Total RNA der Colo 205 Zelllinie (ATCC:CCL-222) eingesetzt.

Die RNA der Zelllinie wurde mit dem radioaktiv markiertem Primer revers transkribiert und auf ein 10%- Polyacrylamidgel aufgetragen. Zur Bestimmung der genauen Bandenlänge wurde ein PCR Fragment von nt 1000 - nt 1362 mit ³⁵S markierten Nukleotiden sequenziert und ebenfalls aufgetragen. Das aus der Elongation des reversen Primers resultierende Fragment von 209 Nukleotiden legt den Transkriptionsstart genau an Position 201 fest. Somit wurde die in Beispiel 6a erhaltene B345-Sequenz in der 5`Region erweitert und ein neues Startkodon auf Position 283 bestimmt. Durch wiederholte Sequenzierungen auch in der 3`Region wurde, im Vergleich zur in SEQ ID NO:1 dargestellten Sequenz, eine zusätzliche Base an Position 2430 gefunden, die zu einer Leserasterverschiebung führt und somit das Stoppcodon von Position 2729 auf 2791 verschiebt. Die in diesem Versuch erhaltene cDNA (SEQ ID NO:3) besitzt einen offenen Leserahmen, der für ein potentielles

Protein mit einer Länge von 836 Aminosäuren (SEQ ID NO:4) kodiert.

Beispiel 7

5 Bioinformatik-Analyse zur Funktion von B345

Die resultierende primäre Aminosäuresequenz von B345 ist in SEQ ID NO:4 gezeigt. Die Analyse des Hydrophilizitäts Plots der Aminosäuresequenz mit der Methode von Kyte und Doolittle (1982) zeigt, dass das B345 Protein zwei

10 charakteristische hydrophobe Domänen aufweist (Aminosäuren Pos. 1 - 29 und 666 - 691), die ein Signalpeptid und eine helikale Transmembrandomäne darstellen (Fig. 6). Diese polarisierte Struktur deutet darauf hin, dass es sich bei B345 um ein integrales
15 Membranprotein handelt. Die Transmembranhelix verbindet einen etwa 666 Aminosäuren langen extrazellulären und einen kurzen (145 Aminosäuren) intrazellulären Teil (siehe Fig. 7).

20 Die extrazelluläre Domäne weist außerdem klare Indizien für die Existenz einer CUB Domäne bei Position 220 - 350 sowie Indizien für 2 mögliche weitere CUB Domänen im Bereich der Aminosäuren 425 - 660 auf. CUB Domänen kommen bei verschiedenen, meist während der Embryonalentwicklung regulierten Proteinen vor. Außerdem
25 sind manchmal bei EGF (Epidermal Growth Factor)-ähnlichen Domänen auch CUB Domänen anzutreffen. Eine kürzlich erschienene Publikation (Gerstein et al., 2000) demonstriert, daß CUB Domänen enthaltende Proteine die am ausgeprägtesten differenziell regulierten

Proteine in *C. elegans* sind. Da Gene, die eine Schlüsselrolle in der Embryonalentwicklung haben, auch analoge Funktionen in Krebs ausführen, kann angenommen werden, dass eine Überexpression von B345 in Zellen eine

5 Veränderung in den Eigenschaften der Substratadhäsion oder der extrazellulären Matrix bewirkt. Das Protein weist außerdem 12 potentielle N- Glycosylierungsstellen auf, die in der vorhergesagten extrazellulären Domäne zu finden sind, was mit der vorhergesagten Orientierung des

10 Proteins übereinstimmt.

Mit einem BLAST hit (E-value: 5.8×10^{-2}) für den Bereich der Aminosäuren von 235 bis 282 von B345 konnte eine Komplement aktivierende Komponente des *RA-reactive factor* (RARF) aus *mus musculus* identifiziert werden. Das

15 Alignment befindet sich innerhalb der CUB Domäne 1 von B345.

Die CUB-Domänen 2 and 3 (Bereich 425-535 und 545-660) weisen marginale Homologien zu dem humanen und Fugu Prokollagen C-Proteinase Enhancer Protein (PCOLCE) auf.

20 Diese Regionen kommen in dem Bereich von PCOLCE vor, welcher ein CUB Domänen Tandem Repeat enthält (E-values: 0.5 (human) und 2.7 (Fugu)). CUB Domänen kommen manchmal in Repeats vor.

Vermutlich bildet das B345-Protein eine β -Sheet

25 Sekundärstruktur, da sich CUB Domänen bekanntlich als β -Sandwich falten.

Die intrazelluläre Domäne (Bereich 691-836) weist keine signifikanten Homologien auf. Der gesamte C-Terminus aligned jedoch mit einem EST (AW063026) von humanen

Eierstockkrebs Zellen (82% Identität über 124 Aminosäuren).

Beispiel 8

5 Bestimmung der genauen Genstruktur von B345

Zunächst wurden Bac Klone in öffentlichen Datenbanken (BLAST search) gesucht, die das B345-Gen enthalten. Die Bac Klone Ac068625 und Ac010170 enthielten einen Großteil des Gens.

Mit intronspannenden Primern wurden Spliceakzeptor und Donorsequenzen in Colo 205 cDNA und genomicischer DNA als Template gesucht. Das PCR Protokoll wurde wie folgt durchgeführt: 1x 95°C 2 Minuten, 35x 95°C 15 Sekunden, 68°C 3 Minuten und dann auf 4°C gehalten. Die PCR wurde auf einem 1,2% Agarosegel analysiert und dabei die Längen der PCR Produkte der 2 Templates mit gleichen Primerkombinationen verglichen. Es stellte sich heraus, dass B345 aus 8 Exons, getrennt von 7 Introns besteht (Fig. 5).

Die chromosomal Lokalisation des Gens wurde mittels

20 Fluoreszenz-in situ-Hybridisierung (FISH) bestimmt.

Dabei wurde die humane, Digoxigenin-markierte B345 Sonde zusammen mit der Biotin-markierten Sonde von B47a2 (Knight et al., 1997), welche sich auf der subtelomerischen Region des Chromosomenarms 3p befindet,

25 mit Metaphase-Chromosomen zweier „normaler“ Individuen hybridisiert (Lichter et al., 1988). Die hybridisierte Digoxigenin-Sonde wurde mittels Anti-Schaf-Dig (Boehringer Mannheim FRG) und Kaninchen Anti-Schaf FITC-markierten Antikörpern detektiert. Die Biotin markierte

Probe dagegen wurde mit Maus Anti-Biotin und Kaninchen-Anti-Maus (TRITC) und anschließende Färbung mit DAPI sichtbar gemacht. Die FISH Resultate zeigen, dass eine Mehrheit der Metaphasen eindeutige Signale an einem oder 5 beiden Chromatiden des Chromosoms 3 in der Region p21-p23 haben. Als Bestätigung der Position diente die Co-lokalisation der B47a2 (TRITC)-Sonde auf dem selben chromosomal Arm.

Tab. 1A

Gewebe	Expression
PBL	-
Lunge	++
Plazenta	+
Dünndarm	+
Leber	-
Niere	++
Milz	-
Thymus	-
Kolon	+
Skelettmuskel	-
Herz	-
Hirn	-

Tab. 1B

Zelllinie	Expression
promyelocytische Leukämie HL60	-
HELA Zellen S3	-
Chronische Myelogene Leukämie K-562	+
Lymphoblastische Leukämie MOLT-4	-
Burkitt's Lymphom (Raji)	-
Kolon Adenokarzinom SW480	+++
Lungen Adenokarzinom A549	+
Melanom G361	-

Tab. 1C

Gewebe	Expression
Speiseröhre Tumor	(+)
Speiseröhre Normal	(+)
Magen Tumor	-
Magen Normal	+
Kolon Tumor	+++
Kolon Normal	++
Mastdarm Tumor	+
Mastdarm Normal	(+)

Tab. 2A

Gewebe	Expression B345/ Actin	Expression B345/ Tubulin
Lunge Adenokarzinom	+	+
Lunge Adenokarzinom	+	+
Lunge Normal	- bis (+)	(+)
Kolon Adenokarzinom	++	++
Kolon Adenokarzinom	+++	+++
Kolon Normal	- bis (+)	+
Mamma IDC	+	+
Brust	-	-
Hodgkin's Lymphom	-	-
Milz	-	-
Testis	-	-

Tab. 2B

Zelllinien und Gewebe	Expression B345 / GAPDH
Kolon Adenokarzinom SW480	+
Kolon Normal (Clonetech)	(+)
Kolon Normal (Invitrogen)	(+)
Lungen Adenokarzinom A549	(+)
Kolon Adenokarzinom Colo 205	+++
PG 102142 Tumor (Colon Ac.)	+++
PG 21900 Tumor (Colon Ac.)	++
PG 7066 Tumor (Colon Ac.)	+++
PG 32389 Tumor (Colon Ac.)	++

Literatur

Abe, K., Rapid isolation of desired sequences from lone linker PCR amplified cDNA mixture: application to
5 identification and recovery of expressed sequences in cloned genomic DNA. *Mamm. Genome* 2, 252-259

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang,
J., Zhang, Z., Miller, W. and Lipman, D. J. Gapped
BLAST and PSI-BLAST: a new generation of protein
10 database search programs. *Nucleic Acids Res.* 25,
3389-3402 (1997).

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe,
K. L. and Sonnhammer, E. L. The Pfam Protein Families
Database. *Nucleic Acids Res.* 28, 263-266 (2000).

15 Bird, A. P. (1986). CpG-rich islands and the function
of DNA methylation. *Nature* 321:209-213.

Bork, P. et al. (1993). The CUB domain. A widespread
module in developmentally regulated proteins. *J.Mol.
Biol.* 231: 539-545

20 Boulianne, G. L., et al., (1984), *Nature* 312: 643-646
Böhm et al., *A.J. of Pathology* 151, 1:63-67, 1997
Brüggemann, M. und Neuberger, M.S., (1996), *Immunol.
Today* 17: 391-397

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M.
25 and Barton, G. J. Jpred: a consensus secondary

structure prediction server. *Bioinformatics* 14,
892-893 (1998).

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik,
A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov,
5 K., Gurskaya, N., Sverdlov, E.D., and Siebert, P.D.
(1996), *Proc. Natl. Acad. Sci. U. S. A.* 93,
6025-6030.

Fields S, Song O (1989) A novel genetic system to
detect protein-protein interactions. *Nature* 20;
10 340(6230):245-6

Gaudi C, Kremer F, Angevin E, Scott V, Triebel F
(1999), *J Immunol* 162:1730-1738

Gerstein, M. and Jansen, R. (2000). The current
excitement in bioinformatics-analysis of whole-genome
15 expression data: how does it relate to protein
structure and function. *Curr. Opin. Struct. Biol.*
10:574-584.

Graziano, R.F., et al., (1995), *J. Immunol.* 155:
4996-5002

20 Griffiths, A.D., et al., (1994), *EMBO J.* 13: 3245-3260
Grosveld, F. and Kollias, G. *Transgenic Animals*,
Academic Press (1992)

Hamburger, A.W. and Salmon, S.E. (1977). Primary
bioassay of human tumor stem cells. *Science*. 197
25 (4302):461-463.

Hesketh, R., (1995), *The oncogene*, Academic Press

Higgins, D. G., Thompson, J. D. and Gibson, T. J. Using CLUSTAL for Multiple Sequence Alignments. *Methods Enzymol.* 266, 383-402 (1996).

5 Hogan KT, Eisinger DP, Cupp SBC, Lekstrom KJ, Deacon DD, Shabanowitz J, Hunt DF, Engelhard VH, Slingluff CL, Ross MM (1998), *Cancer Res* 58:5144-5150

Hofmann, K., Buchner, P., Falquet, L. and Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215-219 (1999).

10 Hubank, M. and Schatz, D.G., (1994), *Nucleic Acids Res.* 22, 5640-5648.

Jakobovits, A., (1995), *Curr. Opin. Biotechnol.* 6: 561-566

15 Kasten, M.B., (1997), *Genetic Instability and Tumorigenesis*, Springer Verlag

20 Knight, S. J., Horsley, S. W., Regan, R., Lawrie, N. M., Maher, E. J., Cardy, D. L., Flint, J., and Kearney, L. (1997). Development and clinical application of an innovative fluorescence *in situ* hybridization technique which detects submicroscopic rearrangements involving telomeres. *Eur. J. Hum. Genet.* 5:1-8.

Köhler, G. und Milstein, C. (1975), *Nature* 265, 495-497

25 Kozak, M (1987), An analysis of 5`noncoding sequences from 99 vertebrates messenger RNA's . *Nuc.Ac.Res.* Vol.15: 8125-8147

Kruif, J., et al., (1995), Proc. Natl. Acad. Sci.
USA 92: 3938-3942

Kyte, J and Doolittle, RF (1982), J. Mol. Biol. 157:
105-132

5 Liaw, L., Skinner, M.P., Raines, E.W., Ross, R.,
Cheresh, D.A., Schwartz, S.M., and Giachelli, C.M..
(1995). The adhesive and migratory effects of
osteopontin are mediated via distinct cell surface
integrins. Role of alpha v beta 3 in smooth muscle
cell migration to osteopontin in vitro.
J.Clin.Invest. 95 (2):713-724.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L.,
and Ward, D. C. (1988). Delineation of individual
human chromosomes in metaphase and interphase cells
15 by in situ suppression hybridization using
recombinant DNA libraries. *Hum. Genet.* 80:224-
234. Mandruzzato S, Brasseur F, Andry G, Boon T, van
der Bruggen P (1997) , *J Exp Med* 186:785-793

McGuinnes, B.T., et al., (1996), *Nature Biotechnol.* 14,
20 1149

Neuberger, M.S., et al., (1984), *Nature* 312: 604-608

Pearson, W. R. and Lipman, D. J. Improved tools for
biological sequence comparison. *Proc. Natl. Acad.*
Sci. USA 85, 2444-2448 (1988).

25 Pusztal et al., 1996, cell proliferation in cancer,
Oxford medical publications

Rauscher, F.J. et al., (1997), Chromosomal translocations and oncogenic transcription factors, Springer Verlag

Riechmann, L., et al., (1988), Nature 332: 323-327

5 Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). "Molecular Cloning: A laboratory Manual" 2nd ed., Cold Spring Laboratory Press, Cold Spring Harbor, NY.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. SMART: A Web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231-234 (2000).

10 Toes, R.E., Hoeben, R.C., Van der Voort, E., Ressing, M.E., Van-der-Eb, A.J. Melief, C.J.M., and Offringa, R. (1997), Proc. Natl. Acad. Sci. U.S.A. 94 (26): 14660-14665

Velculescu, VE, Zhang, L, Vogelstein, B, and Kinzler, KW (1995), Science 270: 484-487

Winter, G., et al., (1994), Annu. Rev. Immunol. 12, 433-455

20 Woelfel, T, Schneider, J, Zum Buschenfelde, Meyer, KH, Rammensee, HG, Rotzschke, O, and Falk, K, (1994), Int. J. Cancer 57: 413-418

Wu, T.C., Guarnieri, F.G., Staveley-O'Carroll, K.F., Viscidi, R.P., Levitsky, H.I., Hedrick, L., Cho, K.R., August, J.T., and Pardoll, D.M. (1995), Proc. Natl. Acad. Sci. U.S.A. 92(25): 11671-11675.

Patentansprüche

1. Tumorassoziiertes Antigen der Bezeichnung B345, dadurch gekennzeichnet, dass es die in SEQ ID NO: 4 definierte Aminosäuresequenz aufweist oder diese als Teilsequenz enthält, oder ein Fragment davon.
2. Isoliertes DNA-Molekül, kodierend für das in Anspruch 1 definierte tumorassoziierte Antigen oder für Fragmente davon.
3. DNA-Molekül nach Anspruch 2, dadurch gekennzeichnet, dass es ein Polynukleotid mit der in SEQ ID NO: 3 dargestellten Sequenz ist oder diese Sequenz enthält oder dass es ein Polynukleotid ist oder dieses enthält, das mit einem Polynukleotid der in SEQ ID NO:3 dargestellten Sequenz unter stringenten Bedingungen hybridisiert, oder ein Fragment davon.
4. Rekombinantes DNA-Molekül, enthaltend ein DNA-Molekül gemäß Anspruch 2 oder 3.
5. Antikörper gegen das in in Anspruch 1 definierte Polypeptid.
6. Antikörper nach Anspruch 5, dadurch gekennzeichnet, dass er monoklonal ist.
7. Antikörper nach Anspruch 5 oder 6 für die Therapie und Diagnose von Krebserkrankungen, die mit der Expression von B345 assoziiert sind.

SEQUENZPROTOKOLL

<110> Boehringer Ingelheim International GmbH

<120> Tumorexprimiertes Polypeptid B345

<130> case12_219

<140>

<141>

<160> 40

<170> PatentIn Ver. 2.1

<210> 1
 <211> 5897
 <212> DNA
 <213> Homo sapiens

<220>
 <221> 5'UTR
 <222> (1)..(214)

<220>
 <221> CDS
 <222> (215)..(2464)

<220>
 <221> 3'UTR
 <222> (2465)..(5897)

<400> 1
 cttgagatat tagaattcgc gactcctgaa ctgcgggtc tctatcgac tgctagggt 60
 tctgctgctg ggtgcggcgc gcctgcccg cgccccagaa gcttttgaga ttgctctgcc 120
 acgagaaaagc aacattacag ttctcataaa gctggggacc ccgactctgc tggcaaaacc 180
 ctgttacatc gtcatttcta aaagacatat aacc atg ttg tcc atc aag tct gga 235
 Met Leu Ser Ile Lys Ser Gly
 1 5

gaa aga ata gtc ttt acc ttt agc tgc cag agt cct gag aat cac ttt 283
 Glu Arg Ile Val Phe Thr Phe Ser Cys Gln Ser Pro Glu Asn His Phe
 10 15 20

gtc ata gag atc cag aaa aat att gac tgt atg tca ggc cca tgt cct 331
 Val Ile Glu Ile Gln Lys Asn Ile Asp Cys Met Ser Gly Pro Cys Pro
 25 30 35

ttt ggg gag gtt cag ctt cag ccc tcg aca tcg ttg ttg cct acc ctc 379
 Phe Gly Glu Val Gln Leu Gln Pro Ser Thr Ser Leu Leu Pro Thr Leu
 40 45 50 55

aac aga act ttc atc tgg gat gtc aaa gct cat aag agc atc ggt tta		427
Asn Arg Thr Phe Ile Trp Asp Val Lys Ala His Lys Ser Ile Gly Leu		
60	65	70
gag ctg cag ttt tcc atc cct cgc ctg agg cag atc ggt ccg ggt gag		475
Glu Leu Gln Phe Ser Ile Pro Arg Leu Arg Gln Ile Gly Pro Gly Glu		
75	80	85
agc tgc cca gac gga gtc act cac tcc atc agc ggc cga atc gat gcc		523
Ser Cys Pro Asp Gly Val Thr His Ser Ile Ser Gly Arg Ile Asp Ala		
90	95	100
acc gtg gtc agg atc gga acc ttc tgc agc aat ggc act gtg tcc cg		571
Thr Val Val Arg Ile Gly Thr Phe Cys Ser Asn Gly Thr Val Ser Arg		
105	110	115
atc aag atg caa gaa gga gtg aaa atg gcc tta cac ctc cca tgg ttc		619
Ile Lys Met Gln Glu Gly Val Lys Met Ala Leu His Leu Pro Trp Phe		
120	125	130
135		
cac ccc aga aat gtc tcc ggc ttc agc att gca aac cgc tca tct ata		667
His Pro Arg Asn Val Ser Gly Phe Ser Ile Ala Asn Arg Ser Ser Ile		
140	145	150
aaa cgt ctg tgc atc atc gag tct gtg ttt gag ggt gaa ggc tca gca		715
Lys Arg Leu Cys Ile Ile Glu Ser Val Phe Glu Gly Glu Gly Ser Ala		
155	160	165
acc ctg atg tct gcc aac tac cca gaa ggc ttc cct gag gat gag ctc		763
Thr Leu Met Ser Ala Asn Tyr Pro Glu Gly Phe Pro Glu Asp Glu Leu		
170	175	180
atg acg tgg cag ttt gtc gtt cct gca cac ctg cgg gcc agc gtc tcc		811
Met Thr Trp Gln Phe Val Val Pro Ala His Leu Arg Ala Ser Val Ser		
185	190	195
ttc ctc aac ttc aac ctc tcc aac tgt gag agg aag gag gag cgg gtt		859
Phe Leu Asn Phe Asn Leu Ser Asn Cys Glu Arg Lys Glu Glu Arg Val		
200	205	210
215		
gaa tac tac atc ccg ggc tcc acc acc aac ccc gag gtg ttc aag ctg		907
Glu Tyr Tyr Ile Pro Gly Ser Thr Thr Asn Pro Glu Val Phe Lys Leu		
220	225	230
gag gac aag cag cct ggg aac atg gcg ggg aac ttc aac ctc tct ctg		955
Glu Asp Lys Gln Pro Gly Asn Met Ala Gly Asn Phe Asn Leu Ser Leu		
235	240	245
caa ggc tgt gac caa gat gcc caa agt cca ggg atc ctc cgg ctg cag		1003
Gln Gly Cys Asp Gln Asp Ala Gln Ser Pro Gly Ile Leu Arg Leu Gln		
250	255	260
ttc caa gtt ttg gtc caa cat cca caa aat gaa agc aat aaa atc tac		1051
Phe Gln Val Leu Val Gln His Pro Gln Asn Glu Ser Asn Lys Ile Tyr		
265	270	275

gtg gtt gac ttg agt aat gag cga gcc atg tca ctc acc atc gag cca		1099
Val Val Asp Leu Ser Asn Glu Arg Ala Met Ser Leu Thr Ile Glu Pro		
280	285	290
295		
cg g ccc gtc aaa cag agc cgc aag ttt gtc cct ggc tgt ttc gtg tgt		1147
Arg Pro Val Lys Gln Ser Arg Lys Phe Val Pro Gly Cys Phe Val Cys		
300	305	310
cta gaa tct cgg acc tgc agt agc aac ctc acc ctg aca tct ggc tcc		1195
Leu Glu Ser Arg Thr Cys Ser Ser Asn Leu Thr Leu Thr Ser Gly Ser		
315	320	325
aaa cac aaa atc tcc ttc ctt tgt gat gat ctg aca cgt ctg tgg atg		1243
Lys His Lys Ile Ser Phe Leu Cys Asp Asp Leu Thr Arg Leu Trp Met		
330	335	340
aat gtg gaa aaa acc ata agc tgc aca gac cac cgg tac tgc caa agg		1291
Asn Val Glu Lys Thr Ile Ser Cys Thr Asp His Arg Tyr Cys Gln Arg		
345	350	355
aaa tcc tac tca ctc cag gtg ccc agt gac atc ctc cac ctg cct gtg		1339
Lys Ser Tyr Ser Leu Gln Val Pro Ser Asp Ile Leu His Leu Pro Val		
360	365	370
375		
gag ctg cat gac ttc tcc tgg aag ctg ctg gtg ccc aag gac agg ctc		1387
Glu Leu His Asp Phe Ser Trp Lys Leu Leu Val Pro Lys Asp Arg Leu		
380	385	390
agc ctg gtg ctg gtg cca gcc cag aag ctg cag cag cat aca cac gag		1435
Ser Leu Val Leu Val Pro Ala Gln Lys Leu Gln Gln His Thr His Glu		
395	400	405
aag ccc tgc aac acc agc ttc agc tac ctc gtg gcc agt gcc ata ccc		1483
Lys Pro Cys Asn Thr Ser Phe Ser Tyr Leu Val Ala Ser Ala Ile Pro		
410	415	420
agc cag gac ctg tac ttc ggc tcc ttc tgc ccg gga ggc tot atc aag		1531
Ser Gln Asp Leu Tyr Phe Gly Ser Phe Cys Pro Gly Gly Ser Ile Lys		
425	430	435
cag atc cag gtg aag cag aac atc tcg gtg acc ctt cgc acc ttt gcc		1579
Gln Ile Gln Val Lys Gln Asn Ile Ser Val Thr Leu Arg Thr Phe Ala		
440	445	450
455		
ccc agc ttccaa aag gag gcc tcc agg cag ggt ctg acg gtg tcc ttt		1627
Pro Ser Phe Gln Gln Glu Ala Ser Arg Gln Gly Leu Thr Val Ser Phe		
460	465	470
ata cct tat ttc aaa gag gaa ggc gtt ttc acg gtg acc cct gac aca		1675
Ile Pro Tyr Phe Lys Glu Glu Gly Val Phe Thr Val Thr Pro Asp Thr		
475	480	485
aaa agc aag gtc tac ctg agg acc ccc aac tgg gac cgg ggc ctg cca		1723
Lys Ser Lys Val Tyr Leu Arg Thr Pro Asn Trp Asp Arg Gly Leu Pro		
490	495	500

tcc ctc acc tct gtg tcc tgg aac atc agc gtg ccc aga gac cag gtg Ser Leu Thr Ser Val Ser Trp Asn Ile Ser Val Pro Arg Asp Gln Val 505 510 515	1771
gcc tgc ctg act ttc ttt aag gag cgg agc ggc gtg gtc tgc cag aca Ala Cys Leu Thr Phe Phe Lys Glu Arg Ser Gly Val Val Cys Gln Thr 520 525 530 535	1819
ggg cgc gca ttc atg atc atc cag gag cag cgg acc cgg gct gag gag Gly Arg Ala Phe Met Ile Ile Gln Glu Arg Thr Arg Ala Glu Glu 540 545 550	1867
atc ttc agc ctg gac gag gat gtg ctc ccc aag cca agc ttc cac cat Ile Phe Ser Leu Asp Glu Asp Val Leu Pro Lys Pro Ser Phe His His 555 560 565	1915
cac agc ttc tgg gtc aac atc tct aac tgc agc ccc acg agc ggc aag His Ser Phe Trp Val Asn Ile Ser Asn Cys Ser Pro Thr Ser Gly Lys 570 575 580	1963
cag cta gac ctg ctc ttc tcg gtg aca ctt acc cca agg act gtg gac Gln Leu Asp Leu Leu Phe Ser Val Thr Leu Thr Pro Arg Thr Val Asp 585 590 595	2011
ttg act gtc atc ctc atc gca gcg gtg gga ggt gga gtc tta ctg ctg Leu Thr Val Ile Leu Ile Ala Ala Val Gly Gly Val Leu Leu Leu 600 605 610 615	2059
tct gcc ctc ggg ctc atc att tgc tgt gtg aaa aag aag aaa aag aag Ser Ala Leu Gly Leu Ile Ile Cys Cys Val Lys Lys Lys Lys Lys 620 625 630	2107
aca aac aag ggc ccc gct gtg ggt atc tac aat ggc aac atc aat act Thr Asn Lys Gly Pro Ala Val Gly Ile Tyr Asn Gly Asn Ile Asn Thr 635 640 645	2155
gag atg cca ggc agc caa aaa agt ttc aga aag ggc gaa agg aca atg Glu Met Pro Gly Ser Gln Lys Ser Phe Arg Lys Gly Glu Arg Thr Met 650 655 660	2203
act ccc atg tgt atg cag tca tcg agg aca cca tgg tat atg ggc atc Thr Pro Met Cys Met Gln Ser Ser Arg Thr Pro Trp Tyr Met Gly Ile 665 670 675	2251
tgc tac agg att cca gcg gct cct tcc tgc agc cag agg tgg aca cct Cys Tyr Arg Ile Pro Ala Ala Pro Ser Cys Ser Gln Arg Trp Thr Pro 680 685 690 695	2299
acc ggc cgt tcc agg gca cca tgg ggg tct gtc ctc cct ccc cac cca Thr Gly Arg Ser Arg Ala Pro Trp Gly Ser Val Leu Pro Pro His Pro 700 705 710	2347
cca tat gct cca ggg ccc caa ctg caa agt tgg cca ctg agg agc cac Pro Tyr Ala Pro Gly Pro Gln Leu Gln Ser Trp Pro Leu Arg Ser His 715 720 725	2395

ctc ctc gct ccc ctc ctg agt ctg aga gtg aac cgt aca cct tct ccc 2443
 Leu Leu Ala Pro Leu Leu Ser Leu Arg Val Asn Arg Thr Pro Ser Pro
 730 735 740

atc cca aca atg ggg atg taa gcagcaagga cacagacatt cccttactga 2494
 Ile Pro Thr Met Gly Met
 745 750

acactcagga gcccataggag ccagcagaat aacttgatcc attccagacg ctttgctgag 2554
 tttcataaag cagggcactg agacacccgt ccgtgttccct aaccagaaat cctaaagaag 2614
 aggaattata cagaaggaac agcaggaggt tttcctggac accgccaact tcacattgct 2674
 cagtggactc attctaaggg caagacattt aaaaatgtga attccaatct ggatacagtc 2734
 atgacagctc atgtgctcct caacttaggc tgtgcggta gccagcctgt aatgagagga 2794
 gagaggcctg agtcacccatg catagggttg cagcaagccc tggattcaga gtgttaaaca 2854
 gaggcttgcc ctcttcagga caacagttcc aattccaagg agcctacctg aggtccctac 2914
 tctcactggg gtccccagga tgaaaacgac aatgtgcctt tttattatta tttatttggt 2974
 ggtcctgtgt tatttaagag atcaaatgtta taaccaccta gctctttca cctgacttag 3034
 taataactca tactaactgg tttggatgcc tgggttgtga cttctactga ccgctagata 3094
 aacgtgtgcc tgccccccag gtgggtggaa taatttacaa tctgtccaac cagaaaagaa 3154
 tgtgtgtgtt tgagcagcat tgacacatat ctgctttgat aagagacttc ctgattctct 3214
 aggtcggttc gtggttatcc cattgtggaa attcatctt aatcccatttgc ttctatagtc 3274
 ctagcaataa gagaatttc ctcaagttc catgtgcggc tctcctagct gcagcaatac 3334
 tttgacattt aaagagaaat ttagagaata ttctcatcct ctaaaaatgt taaaatataat 3394
 accaaacagt ggccccctgc attagtttc tggtgccact gcaacccatt acttggtagc 3454
 ttaaaaacaa cacattagct tatagtcctg gggatcagaa ttccaaaatg gatgtccctg 3514
 aatgaaaatc aagggtgtcag cagagctgtg ctccttcga aggctctagg gagaagccgg 3574
 ttccttgcca tttcaagctt ctagaggctg gctgcattcc caggctccag tggctggta 3634
 agctttctc acatggcattc actgtgacac tggccctccc acttccctct ttgacttaca 3694
 aagcccacca ggaagatcca ggataatctc tccatctaaa gatccttcat catcctggaa 3754
 gagccttttgc ccatgcaaga caacatagcc acaggtgggg attaggacca ggacatcttt 3814
 ggggtgtgtt tattctgcctt accacacctt cctgccacbg actcccacag gagaggctac 3874
 aaaatgatct ggccgcacagg gatgtttgtt ttagcttgcg gactctaaca cttaaaaaaaa 3934
 ccccagatca gaagatctgg ccatgctgg gctcacattc tcacctagca acaactggct 3994

ggagactgggc accagctctg cctttagaag gggtgtccac ttcaccaggc caccacagcc 4054
cacactacgc cctatcaatt cccacaatga ggctaagtgt ttgtttctac tgatcaatgc 4114
ccctgcaggt tgcattttt gtaatgaaaa agaaagactg ggattaatct ctaatcaggt 4174
gagtagacca tgagaccaa gtgtgctcac attaccctt ttctttttt tcttttctt 4234
tttctttttt ttttaatgt gagacaggat ctcattctgt tgcctaggct ggagtgcagt 4294
ggcgcaatct cggctcaactg caacctctgc ctccctggct caagcaattc tcccacacta 4354
gcctcccaaa tagctggat cactggcaca aaccaccatg cccagctaatt tttgtatttt 4414
ttgttagagac agggtttcac catgttgcac aggctggtct caacccctgc ggctcaagca 4474
atcctcctgc ctccggcctcc caaagtgcgtg ggattacaga tgtgagccac cgcatccagc 4534
ccccacccct catttataacc aattacctgc ccagtaactg tggacttttgc ttccctcacc 4594
cctgctctga tctggaaagga gagggattat gttatagctt gtcagcacag tcccaagttc 4654
aatatttctg cggcaaaaac ttccctcaaa aaataaatgt acttcattgtt attcaatgaa 4714
ttcaccttgg aaatgcacccg cctcaacttg ttcacatggc ataaatgaaa ggaattttat 4774
agtctcctaa atggcgtgta ctgcaagacc tcttgaacac ttccagagg ataggatatt 4834
taagtcatgc ccttggcggt gcctatggca ccttccctt ctgaaagtct ggccctgcc 4894
cagtgaccct tggccttggc agccgagatg ctgaccctgc ataaaggggcc aaaggagggc 4954
tgccggcttcc ttccctcaact gaagagccct tatttgaatt cactgtgtgg agccctagcc 5014
ctccattctc gacattcccc aacctcccag ccccttccaa gcaggactag gtccctgca 5074
ttccacccaa ggtgggatttgc ccttcctta ggctggctac ttgtcaccat caccgacatc 5134
actgttgctt gcaaggacac cacgtggcca ttttccttca actgagggct caaaactcct 5194
ggacaagtttgc tggccttgc agaccagtat ttcctggagm tgcctcag tgaagggggcc 5254
cagcctgagg aaccctggctt ctttcttta aagcccagggc cccacttaca taaaacattt 5314
cagggtcactt gaaaacagtgc aagtgcatttgc tgccttgc tgccttgc tgccttgc 5374
ctcatccacg tggatgccttgc tgccttgc tgccttgc tgccttgc tgccttgc 5434
atgnntgtggcatttgc acatggccat tgccttgc tgccttgc tgccttgc tgccttgc 5494
aagcccacccg gacatggccat tggtaaagggt tagcagactg gtgtgtgtgg atctgcagtg 5554
cttcacttgc aataatttttgc tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc 5614
tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc 5674
tacttgccttgc tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc tgccttgc 5734

gcatgggctc agcagatgct gtcttaattt gtggatgata cagaaagcca ggcttgga 5794
 tacaagttct ttccctttca tttgatgccg tgcactgtgt gaagcagatg ttttgtccg 5854
 gaaataaaaa taatagtctt ggagtctcgc caaaaaaaaaa aag 5897

<210> 2
 <211> 749
 <212> PRT
 <213> Homo sapiens

<400> 2															
Met	Leu	Ser	Ile	Lys	Ser	Gly	Glu	Arg	Ile	Val	Phe	Thr	Phe	Ser	Cys
1				5					10					15	
Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn Ile Asp															
		20				25							30		
Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln Pro Ser															
		35			40							45			
Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp Val Lys															
		50			55					60					
Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe Ser Ile Pro Arg Leu															
		65			70				75			80			
Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp Gly Val Thr His Ser															
		85				90					95				
Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg Ile Gly Thr Phe Cys															
		100			105					110					
Ser Asn Gly Thr Val Ser Arg Ile Lys Met Gln Glu Gly Val Lys Met															
		115			120				125						
Ala Leu His Leu Pro Trp Phe His Pro Arg Asn Val Ser Gly Phe Ser															
		130			135				140						
Ile Ala Asn Arg Ser Ser Ile Lys Arg Leu Cys Ile Ile Glu Ser Val															
		145			150				155			160			
Phe Glu Gly Glu Gly Ser Ala Thr Leu Met Ser Ala Asn Tyr Pro Glu															
		165				170				175					
Gly Phe Pro Glu Asp Glu Leu Met Thr Trp Gln Phe Val Val Pro Ala															
		180			185				190						
His Leu Arg Ala Ser Val Ser Phe Leu Asn Phe Asn Leu Ser Asn Cys															
		195			200				205						
Glu Arg Lys Glu Glu Arg Val Glu Tyr Tyr Ile Pro Gly Ser Thr Thr															
		210			215				220						
Asn Pro Glu Val Phe Lys Leu Glu Asp Lys Gln Pro Gly Asn Met Ala															
		225			230				235			240			

Gly Asn Phe Asn Leu Ser Leu Gln Gly Cys Asp Gln Asp Ala Gln Ser
 245 250 255

Pro Gly Ile Leu Arg Leu Gln Phe Gln Val Leu Val Gln His Pro Gln
 260 265 270

Asn Glu Ser Asn Lys Ile Tyr Val Val Asp Leu Ser Asn Glu Arg Ala
 275 280 285

Met Ser Leu Thr Ile Glu Pro Arg Pro Val Lys Gln Ser Arg Lys Phe
 290 295 300

Val Pro Gly Cys Phe Val Cys Leu Glu Ser Arg Thr Cys Ser Ser Asn
 305 310 315 320

Leu Thr Leu Thr Ser Gly Ser Lys His Lys Ile Ser Phe Leu Cys Asp
 325 330 335

Asp Leu Thr Arg Leu Trp Met Asn Val Glu Lys Thr Ile Ser Cys Thr
 340 345 350

Asp His Arg Tyr Cys Gln Arg Lys Ser Tyr Ser Leu Gln Val Pro Ser
 355 360 365

Asp Ile Leu His Leu Pro Val Glu Leu His Asp Phe Ser Trp Lys Leu
 370 375 380

Leu Val Pro Lys Asp Arg Leu Ser Leu Val Leu Val Pro Ala Gln Lys
 385 390 395 400

Leu Gln Gln His Thr His Glu Lys Pro Cys Asn Thr Ser Phe Ser Tyr
 405 410 415

Leu Val Ala Ser Ala Ile Pro Ser Gln Asp Leu Tyr Phe Gly Ser Phe
 420 425 430

Cys Pro Gly Gly Ser Ile Lys Gln Ile Gln Val Lys Gln Asn Ile Ser
 435 440 445

Val Thr Leu Arg Thr Phe Ala Pro Ser Phe Gln Gln Glu Ala Ser Arg
 450 455 460

Gln Gly Leu Thr Val Ser Phe Ile Pro Tyr Phe Lys Glu Glu Gly Val
 465 470 475 480

Phe Thr Val Thr Pro Asp Thr Lys Ser Lys Val Tyr Leu Arg Thr Pro
 485 490 495

Asn Trp Asp Arg Gly Leu Pro Ser Leu Thr Ser Val Ser Trp Asn Ile
 500 505 510

Ser Val Pro Arg Asp Gln Val Ala Cys Leu Thr Phe Phe Lys Glu Arg
 515 520 525

Ser Gly Val Val Cys Gln Thr Gly Arg Ala Phe Met Ile Ile Gln Glu
 530 535 540

Gln Arg Thr Arg Ala Glu Glu Ile Phe Ser Leu Asp Glu Asp Val Leu
 545 550 555 560

Pro Lys Pro Ser Phe His His Ser Phe Trp Val Asn Ile Ser Asn
 565 570 575

Cys Ser Pro Thr Ser Gly Lys Gln Leu Asp Leu Leu Phe Ser Val Thr
 580 585 590

Leu Thr Pro Arg Thr Val Asp Leu Thr Val Ile Leu Ile Ala Ala Val
 595 600 605

Gly Gly Gly Val Leu Leu Ser Ala Leu Gly Leu Ile Ile Cys Cys
 610 615 620

Val Lys Lys Lys Lys Lys Thr Asn Lys Gly Pro Ala Val Gly Ile
 625 630 635 640

Tyr Asn Gly Asn Ile Asn Thr Glu Met Pro Gly Ser Gln Lys Ser Phe
 645 650 655

Arg Lys Gly Glu Arg Thr Met Thr Pro Met Cys Met Gln Ser Ser Arg
 660 665 670

Thr Pro Trp Tyr Met Gly Ile Cys Tyr Arg Ile Pro Ala Ala Pro Ser
 675 680 685

Cys Ser Gln Arg Trp Thr Pro Thr Gly Arg Ser Arg Ala Pro Trp Gly
 690 695 700

Ser Val Leu Pro Pro His Pro Pro Tyr Ala Pro Gly Pro Gln Leu Gln
 705 710 715 720

Ser Trp Pro Leu Arg Ser His Leu Leu Ala Pro Leu Leu Ser Leu Arg
 725 730 735

Val Asn Arg Thr Pro Ser Pro Ile Pro Thr Met Gly Met
 740 745

```

<210> 3
<211> 6163
<212> DNA
<213> Homo sapiens

<220>
<221> 5'UTR
<222> (1)..(282)

<220>
<221> GC_signal
<222> (147)..(157)

<220>
<221> misc_feature
<222> (201)..(209)
<223> cap signal; Transkriptionsstart
  
```

<220>
 <221> 3'UTR
 <222> (2794)..(6163)

<220>
 <221> 3'UTR
 <222> (2794)..(6163)

<220>
 <221> CDS
 <222> (283)..(2793)

<400> 3
 ccaacccgc aatggggagt agtagggacc cagcaacccg gtgccggag ccctgcaccc 60
 tgggagggag aggccgtcgc tgaggcagga agaggaggag gagagagagg agggacgcac 120
 cgggtcagct cgcgatcctg ctgcgcaggg cggggctcgg gccggtccgc cgcgcgcag 180
 gtgagtgagc cagggcggag cgcaagtcgc ccgggcttgg gcgcctgggg cgcggctcc 240
 ccaccgtcgt ttccccacc gaggccgagg cgtcccggag tc atg gcc ggc ctg 294
 Met Ala Gly Leu
 1

aac tgc ggg gtc tct atc gca ctg cta ggg gtt ctg ctg ctg ggt gcg 342
 Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu Leu Leu Gly Ala
 5 10 15 20

gcg cgc ctg ccg cgc ggg gca gaa gct ttt gag att gct ctg cca cga 390
 Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile Ala Leu Pro Arg
 25 30 35

gaa agc aac att aca gtt ctc ata aag ctg ggg acc ccg act ctg ctg 438
 Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr Pro Thr Leu Leu
 40 45 50

gca aaa ccc tgt tac atc gtc att tct aaa aga cat ata acc atg ttg 486
 Ala Lys Pro Cys Tyr Ile Val Ile Ser Lys Arg His Ile Thr Met Leu
 55 60 65

tcc atc aag tct gga gaa aga ata gtc ttt acc ttt agc tgc cag agt 534
 Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe Ser Cys Gln Ser
 70 75 80

cct gag aat cac ttt gtc ata gag atc cag aaa aat att gac tgt atg 582
 Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn Ile Asp Cys Met
 85 90 95 100

tca ggc cca tgt cct ttt ggg gag gtt cag ctt cag ccc tcg aca tcg 630
 Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln Pro Ser Thr Ser
 105 110 115

ttg ttg cct acc ctc aac aga act ttc atc tgg gat gtc aaa gct cat 678
 Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp Val Lys Ala His
 120 125 130

aag	agc	atc	ggt	tta	gag	ctg	cag	ttt	tcc	atc	cct	cgc	ctg	agg	cag		726
Lys	Ser	Ile	Gly	Leu	Glu	Leu	Gln	Phe	Ser	Ile	Pro	Arg	Leu	Arg	Gln		
135									140						145		
atc	ggt	ccg	ggt	gag	agc	tgc	cca	gac	gga	gtc	act	cac	tcc	atc	agc		774
Ile	Gly	Pro	Gly	Glu	Ser	Cys	Pro	Asp	Gly	Val	Thr	His	Ser	Ile	Ser		
150									155						160		
ggc	cga	atc	gat	gcc	acc	gtg	gtc	agg	atc	gga	acc	ttc	tgc	agc	aat		822
Gly	Arg	Ile	Asp	Ala	Thr	Val	Val	Arg	Ile	Gly	Thr	Phe	Cys	Ser	Asn		
165									170						175		180
ggc	act	gtg	tcc	cg	atc	aag	atg	caa	gaa	gga	gtg	aaa	atg	gcc	tta		870
Gly	Thr	Val	Ser	Arg	Ile	Lys	Met	Gln	Glu	Gly	Val	Lys	Met	Ala	Leu		
185									190						195		
cac	ctc	cca	tgg	ttc	cac	ccc	aga	aat	gtc	tcc	ggc	ttc	agc	att	gca		918
His	Leu	Pro	Trp	Phe	His	Pro	Arg	Asn	Val	Ser	Gly	Phe	Ser	Ile	Ala		
200									205						210		
aac	cgc	tca	tct	ata	aaa	cgt	ctg	tgc	atc	atc	gag	tct	gtg	ttt	gag		966
Asn	Arg	Ser	Ser	Ile	Lys	Arg	Ley	Cys	Ile	Ile	Glu	Ser	Val	Phe	Glu		
215									220						225		
gg	gaa	ggc	tca	gca	acc	ctg	atg	tct	gcc	aac	tac	cca	gaa	ggc	ttc		1014
Gly	Glu	Gly	Ser	Ala	Thr	Leu	Met	Ser	Ala	Asn	Tyr	Pro	Glu	Gly	Phe		
230									235						240		
cct	gag	gat	gag	ctc	atg	acg	tgg	cag	ttt	gtc	gtt	cct	gca	cac	ctg		1062
Pro	Glu	Asp	Glu	Leu	Met	Thr	Trp	Gln	Phe	Val	Val	Pro	Ala	His	Leu		
245									250						255		260
cg	gg	cc	agc	gtc	tcc	ttc	ctc	aac	ttc	aac	ctc	tcc	aac	tgt	gag	agg	1110
Arg	Ala	Ser	Val	Ser	Phe	Leu	Asn	Phe	Asn	Leu	Ser	Asn	Cys	Glu	Arg		
265									270						275		
aag	gag	gag	cg	gtt	gaa	tac	tac	atc	ccg	ggc	tcc	acc	acc	aac	ccc		1158
Lys	Glu	Glu	Arg	Val	Glu	Tyr	Tyr	Ile	Pro	Gly	Ser	Thr	Thr	Asn	Pro		
280									285						290		
gag	gt	ttc	aag	ctg	gag	gac	aag	cag	cct	ggg	aac	atg	g	ggg	aa		1206
Glu	Val	Phe	Lys	Leu	Glu	Asp	Lys	Gln	Pro	Gly	Asn	Met	Ala	Gly	Asn		
295									300						305		
ttc	aa	ctc	tct	ctg	caa	ggc	tgt	gac	caa	gat	gcc	caa	agt	cca	gg		1254
Phe	Asn	Leu	Ser	Ile	Gln	Gly	Cys	Asp	Gln	Asp	Ala	Gln	Ser	Pro	Gly		
310									315						320		
atc	ctc	cg	ctg	cag	ttc	caa	gtt	ttg	gtc	caa	cat	cca	caa	aat	gaa		1302
Ile	Leu	Arg	Leu	Gln	Phe	Gln	Val	Leu	Val	Gln	His	Pro	Gln	Asn	Glu		
325									330						335		340
agc	aat	aaa	atc	tac	gt	tt	gac	tt	gt	agt	aat	gag	cga	gcc	atg	tca	1350
Ser	Asn	Lys	Ile	Tyr	Val	Val	Asp	Leu	Ser	Asn	Glu	Arg	Ala	Met	Ser		
345									350						355		

ctc acc atc gag cca cg ^g ccc gtc aaa cag agc cgc aag ttt gtc cct Leu Thr Ile Glu Pro Arg Pro Val Lys Gln Ser Arg Lys Phe Val Pro 360 365 370	1398
ggc tgt ttc gtg tgt cta gaa tct cg ^g acc tgc agt agc aac ctc acc Gly Cys Phe Val Cys Leu Glu Ser Arg Thr Cys Ser Ser Asn Leu Thr 375 380 385	1446
ctg aca tct ggc tcc aaa cac aaa atc tcc ttc ctt tgt gat gat ctg Leu Thr Ser Gly Ser Lys His Lys Ile Ser Phe Leu Cys Asp Asp Leu 390 395 400	1494
aca cgt ctg tgg atg aat gtg gaa aaa acc ata agc tgc aca gac cac Thr Arg Leu Trp Met Asn Val Glu Lys Thr Ile Ser Cys Thr Asp His 405 410 415 420	1542
cg ^g tac tgc caa agg aaa tcc tac tca ctc cag gtg ccc agt gac atc Arg Tyr Cys Gln Arg Lys Ser Tyr Ser Leu Gln Val Pro Ser Asp Ile 425 430 435	1590
ctc cac ctg cct gtg gag ctg cat gac ttc tcc tgg aag ctg ctg gtg Leu His Leu Pro Val Glu Leu His Asp Phe Ser Trp Lys Leu Leu Val 440 445 450	1638
ccc aag gac agg ctc agc ctg gtg ctg gtg cca gcc cag aag ctg cag Pro Lys Asp Arg Leu Ser Leu Val Leu Val Pro Ala Gln Lys Leu Gln 455 460 465	1686
cag cat aca cac gag aag ccc tgc aac acc agc ttc agc tac ctc gtg Gln His Thr His Glu Lys Pro Cys Asn Thr Ser Phe Ser Tyr Leu Val 470 475 480	1734
gcc agt gcc ata ccc agc cag gac ctg tac ttc ggc tcc ttc tgc ccg Ala Ser Ala Ile Pro Ser Gln Asp Leu Tyr Phe Gly Ser Phe Cys Pro 485 490 495 500	1782
gga ggc tct atc aag cag atc cag gtg aag cag aac atc tcg gtg acc Gly Gly Ser Ile Lys Gln Ile Gln Val Lys Gln Asn Ile Ser Val Thr 505 510 515	1830
ctt cgc acc ttt gcc ccc agc ttc caa caa gag gcc tcc agg cag ggt Leu Arg Thr Phe Ala Pro Ser Phe Gln Gln Glu Ala Ser Arg Gln Gly 520 525 530	1878
ctg acg gtg tcc ttt ata cct tat ttc aaa gag gaa ggc gtt ttc acg Leu Thr Val Ser Phe Ile Pro Tyr Phe Lys Glu Glu Gly Val Phe Thr 535 540 545	1926
gtg acc cct gac aca aaa agc aag gtc tac ctg agg acc ccc aac tgg Val Thr Pro Asp Thr Lys Ser Lys Val Tyr Leu Arg Thr Pro Asn Trp 550 555 560	1974
gac cg ^g ggc ctg cca tcc ctc acc tct gtg tcc tgg aac atc agc gtg Asp Arg Gly Leu Pro Ser Leu Thr Ser Val Ser Trp Asn Ile Ser Val 565 570 575 580	2022

ccc aga gac cag gtg gcc tgc ctg act ttc ttt aag gag cgg agc ggc Pro Arg Asp Gln Val Ala Cys Leu Thr Phe Phe Lys Glu Arg Ser Gly 585 590 595	2070
gtg gtc tgc cag aca ggg cgc gca ttc atg atc atc cag gag cag cgg Val Val Cys Gln Thr Gly Arg Ala Phe Met Ile Ile Gln Glu Gln Arg 600 605 610	2118
acc cggt gct gag gag atc ttc agc ctg gac gag gat gtg ctc ccc aag Thr Arg Ala Glu Glu Ile Phe Ser Leu Asp Glu Asp Val Leu Pro Lys 615 620 625	2166
cca agc ttc cac cat cac agc ttc tgg gtc aac atc tct aac tgc agc Pro Ser Phe His His Ser Phe Trp Val Asn Ile Ser Asn Cys Ser 630 635 640	2214
ccc acg agc ggc aag cag cta gac ctg ctc ttc tcg gtg aca ctt acc Pro Thr Ser Gly Lys Gln Leu Asp Leu Leu Phe Ser Val Thr Leu Thr 645 650 655 660	2262
cca agg act gtg gac ttg act gtc atc ctc atc gca gcg gtg gga ggt Pro Arg Thr Val Asp Leu Thr Val Ile Leu Ile Ala Ala Val Gly Gly 665 670 675	2310
gga gtc tta ctg ctg tct gcc ctc ggg ctc atc att tgc tgt gtg aaa Gly Val Leu Leu Ser Ala Leu Gly Leu Ile Ile Cys Cys Val Lys 680 685 690	2358
aag aag aaa aag aag aca aac aag ggc ccc gct gtg ggt atc tac aat Lys Lys Lys Lys Thr Asn Lys Gly Pro Ala Val Gly Ile Tyr Asn 695 700 705	2406
ggc aac atc aat act gag atg ccg agg cag cca aaa aag ttt cag aaa Gly Asn Ile Asn Thr Glu Met Pro Arg Gln Pro Lys Lys Phe Gln Lys 710 715 720	2454
ggg cga aag gac aat gac tcc cat gtg tat gca gtc atc gag gac acc Gly Arg Lys Asp Asn Asp Ser His Val Tyr Ala Val Ile Glu Asp Thr 725 730 735 740	2502
atg gta tat ggg cat ctg cta cag gat tcc agc ggc tcc ttc ctg cag Met Val Tyr Gly His Leu Leu Gln Asp Ser Ser Gly Ser Phe Leu Gln 745 750 755	2550
cca gag gtg gac acc tac ccg ccg ttc cag ggc acc atg ggg gtc tgt Pro Glu Val Asp Thr Tyr Arg Pro Phe Gln Gly Thr Met Gly Val Cys 760 765 770	2598
cct ccc tcc cca ccc acc ata tgc tcc agg gcc cca act gca aag ttg Pro Pro Ser Pro Pro Thr Ile Cys Ser Arg Ala Pro Thr Ala Lys Leu 775 780 785	2646
gcc act gag gag cca cct cct cgc tcc cct cct gag tct gag agt gaa Ala Thr Glu Glu Pro Pro Arg Ser Pro Pro Glu Ser Glu Ser Glu 790 795 800	2694

ccg tac acc ttc tcc cat ccc aac aat ggg gat gta agc agc aag gac 2742
 Pro Tyr Thr Phe Ser His Pro Asn Asn Gly Asp Val Ser Ser Lys Asp
 805 810 815 820

aca gac att ccc tta ctg aac act cag gag ccc atg gag cca gca gaa 2790
 Thr Asp Ile Pro Leu Leu Asn Thr Gln Glu Pro Met Glu Pro Ala Glu
 825 830 835

taa ctgtatccat tccagacgct ttgctgagg ttataaagca gggcactgag 2843

acaccgtcc gtgttcctaa ccagaaatcc taaaagaagag gaattataca gaaggaacag 2903
 caggaggttt tcctggacac cgccaacttc acattgctca gtggactcat tctaagggca 2963
 agacattgaa aatgatgaat tccaatctgg atacagtcat gacagctcat gtgctcctca 3023
 acttaggctg tgcggttagc cagcctgtaa tgagaggaga gaggcctgag tcacctagca 3083
 tagggttgca gcaagccctg gattcagagt gttaaacaga ggcttgccct cttcaggaca 3143
 acagttccaa ttccaaggag cctacctgag gtccctactc tcactgggt ccccaggatg 3203
 aaaacgacaa tgtgccttt tattattatt tatttgtgg tcctgtgtta tttaagagat 3263
 caaatgtata accacctagc tctttcacc tgacttagta ataactcata ctaactgggt 3323
 tggatgcctg gggtgtgact tctactgacc gctagataaa cgtgtgcctg tccccaggt 3383
 ggtggaaata attacaatc tgtccaacca gaaaagaatg tgtgtgtttg agcagcattg 3443
 acacatatct gcttgataa gagacttcct gattctctag gtccgttcgt ggttatccca 3503
 ttgtggaaat tcatcttcaa tcccattgtc ctatagtcct agcaataaga gaaatttcct 3563
 caagtttcca tgtgcgggtc tcctagctgc agcaatactt tgacattaa agagaaattt 3623
 agagaatatt ctcatcctct aaaaatgttt aaatatatac caaacagtgg ccccctgcat 3683
 tagtttctg ttgccactgc aaccattac ttggtagctt aaaaacaaca cattagctt 3743
 tagcctggg gatcagaatt caaaaatgga tgtccctgaa tgaaaatcaa ggtgtcagca 3803
 gagctgtgct ccttctgaag gctctaggaa gaagccggtt ctttgccatt tcaagcttct 3863
 agaggctggc tgcattccca ggctccagtg gctggtaag ctttctcac atggcatcac 3923
 tgtgacactg gccctccac ttccctctt gacttacaaa gcccaccagg aagatccagg 3983
 ataatctctc catctaaaga tccttcatca tcctggaaaga gcctttgcc atgcaagaca 4043
 acatagccac aggtggggat taggaccagg acatcttgg ggtgtgtta ttctgcctac 4103
 cacacattcc tgcactgac tcccacagga gaggctacaa aatgatctgg cgcacaggg 4163
 tgggggttt agcttgcggaa ctctaacact taaaaaaaaacc ccagatcaga agatctggcc 4223
 atgctggggc tcacattctc acctagcaac aactggctgg agctggcac cagctctgcc 4283

tttagaaggg gtgtccactt caccaggta ccacagccc cactacgccc tatcacttcc 4343
cacaatgagg ctaagtgttt gtttctactg atcaatgccc ctgcagggtt catttattgt 4403
aatgaaaaag aaagactggg attaatctct aatcaggtga gtagaccatg agaccaatgt 4463
gtgctcacat taccctttt ctttttttc tttttctttt tctttttttt tttaatgtga 4523
gacaggatct cattctgttg cctaggctgg agtgcagtgg cgcaatctcg gctcaactgca 4583
acctctgcct cctgggctca agcaattctc ccacctcagc ctcccaaata gctgggatca 4643
ctggcacaaa ccaccatgcc cagctaattt tgtatTTTTT gtagagacag ggTTTcacca 4703
tgttgcccag gctggtctca acctcctggg ctcaagcaat cctcctgcct cggcctccca 4763
aagtgctggg attacagatg tgagccaccc catccagccc cacaccctca tttataccaa 4823
ttacctgccc agtaactgtg gactttgct tcctcacccc tgctctgatc tggaaggaga 4883
gggattatgt tatagcttgt cagcacagtc ccaagttcaa tatttctgag gcaaaaactt 4943
ccttcaaaaa ataaatgtac ttcatgttat tcaatgaatt cacctggaa atgcacccgcc 5003
tcaactgtt cacatggcat aaatgaaagg aattttatag tctcctaaat ggcgtgtact 5063
gcaagacctc ttgaacactt tccagaggat aggatattta agtcatgccc ttggcgttgc 5123
ctatggcacc ttcccttct gaaagtctgg ttcctgccc gtgacccttg gccttgtag 5183
ccgagatgct gaccctgcat aaagggccaa aggagggctg cggcttcctt ccctcactga 5243
agagccctta ttgaattca ctgtgtggag ccctagccct ccattctcga cattccccaa 5303
cctcccagcc cttccaagc aggactaggt gccctgcatt ccacccaagg tgggattggc 5363
cttccttagg ctggctactt gtcaccatca ccgacatcac tggcctgc aaggacacca 5423
cgtggccatt ttcccaac tgagggctca aaactcctgg acaagttgct ggctcctgag 5483
accagtattt cctggagctg tgcctcagtg aaggggccc gcctgaggaa ccctggctct 5543
tttctttaaa gcccaggccc cacttacata aaacattca gggtaactgg aaacagtgaa 5603
gtgccatttg ttgaaggcta ctgcattgcca gcccactgct catccacgtg gtctgcccatt 5663
cctacgagga aggccagcgc atgcaggact ggtctctaatt gctgtgtca ttgcacagaa 5723
ggaaaggta tcaaggaaga gtcaactggg acaagcacaa gcccacccggc catggccttg 5783
gtaaaggta gcagactggt gtgtgtggat ctgcagtgtc tcactggaaa taatTTTattc 5843
attgcagata cttttaggt ggcattttat tcatttcctg tgctttaat aaacaaatgt 5903
accaaaaaaaaac aagtatcaag ctgttaagt gttcggcta cttgtcccct ggTTcagtag 5963
aggccccggc ttccctggat ttgactgtga caggctcagc atgggctcag cagatgctgt 6023

cttaatttgt ggatgataca gaaagccagg ctttggata caagttctt cctcttcatt 6083
tgatgccgtg cactgtgtga agcagatgtt tttgtccgga aataaaaaata atagtcttgg 6143
agtctcgcca aaaaaaaaaaa 6163

<210> 4
<211> 836
<212> PRT
<213> *Homo sapiens*

<400> 4
Met Ala Gly Leu Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu
1 5 10 15

Leu Leu Gly Ala Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile
20 25 30

Ala Leu Pro Arg Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr
35 40 45

Ile Thr Met Leu Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe
65 70 75 80

Ser Cys Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn
85 90 95

Ile Asp Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln
100 105 110

Pro Ser Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp
115 120 125

Val Lys Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe Ser Ile Pro
 130 135 140

Arg Leu Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp Gly Val Thr
145 150 155 160

His Ser Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg Ile Gly Thr
165 / 170 175

Phe Cys Ser Asn Gly Thr Val Ser Arg Ile Lys Met Gln Glu Gly Val
180 | 185 190

Lys Met Ala Leu His Leu Pro Trp Phe His Pro Arg Asn Val Ser Gly
195 200 205

Phe Ser Ile Ala Asn Arg Ser Ser Ile Lys Arg Leu Cys Ile Ile Glu
210 215 220

Ser Val Phe Glu Gly Glu Gly Ser Ala Thr Leu Met Ser Ala Asn Tyr
 225 230 235 240

 Pro Glu Gly Phe Pro Glu Asp Glu Leu Met Thr Trp Gln Phe Val Val
 245 250 255

 Pro Ala His Leu Arg Ala Ser Val Ser Phe Leu Asn Phe Asn Leu Ser
 260 265 270

 Asn Cys Glu Arg Lys Glu Glu Arg Val Glu Tyr Tyr Ile Pro Gly Ser
 275 280 285

 Thr Thr Asn Pro Glu Val Phe Lys Leu Glu Asp Lys Gln Pro Gly Asn
 290 295 300

 Met Ala Gly Asn Phe Asn Leu Ser Leu Gln Gly Cys Asp Gln Asp Ala
 305 310 315 320

 Gln Ser Pro Gly Ile Leu Arg Leu Gln Phe Gln Val Leu Val Gln His
 325 330 335

 Pro Gln Asn Glu Ser Asn Lys Ile Tyr Val Val Asp Leu Ser Asn Glu
 340 345 350

 Arg Ala Met Ser Leu Thr Ile Glu Pro Arg Pro Val Lys Gln Ser Arg
 355 360 365

 Lys Phe Val Pro Gly Cys Phe Val Cys Leu Glu Ser Arg Thr Cys Ser
 370 375 380

 Ser Asn Leu Thr Leu Thr Ser Gly Ser Lys His Lys Ile Ser Phe Leu
 385 390 395 400

 Cys Asp Asp Leu Thr Arg Leu Trp Met Asn Val Glu Lys Thr Ile Ser
 405 410 415

 Cys Thr Asp His Arg Tyr Cys Gln Arg Lys Ser Tyr Ser Leu Gln Val
 420 425 430

 Pro Ser Asp Ile Leu His Leu Pro Val Glu Leu His Asp Phe Ser Trp
 435 440 445

 Lys Leu Leu Val Pro Lys Asp Arg Leu Ser Leu Val Leu Val Pro Ala
 450 455 460

 Gln Lys Leu Gln Gln His Thr His Glu Lys Pro Cys Asn Thr Ser Phe
 465 470 475 480

 Ser Tyr Leu Val Ala Ser Ala Ile Pro Ser Gln Asp Leu Tyr Phe Gly
 485 490 495

 Ser Phe Cys Pro Gly Gly Ser Ile Lys Gln Ile Gln Val Lys Gln Asn
 500 505 510

 Ile Ser Val Thr Leu Arg Thr Phe Ala Pro Ser Phe Gln Gln Glu Ala
 515 520 525

Ser Arg Gln Gly Leu Thr Val Ser Phe Ile Pro Tyr Phe Lys Glu Glu
 530 535 540
 Gly Val Phe Thr Val Thr Pro Asp Thr Lys Ser Lys Val Tyr Leu Arg
 545 550 555 560
 Thr Pro Asn Trp Asp Arg Gly Leu Pro Ser Leu Thr Ser Val Ser Trp
 565 570 575
 Asn Ile Ser Val Pro Arg Asp Gln Val Ala Cys Leu Thr Phe Phe Lys
 580 585 590
 Glu Arg Ser Gly Val Val Cys Gln Thr Gly Arg Ala Phe Met Ile Ile
 595 600 605
 Gln Glu Gln Arg Thr Arg Ala Glu Glu Ile Phe Ser Leu Asp Glu Asp
 610 615 620
 Val Leu Pro Lys Pro Ser Phe His His His Ser Phe Trp Val Asn Ile
 625 630 635 640
 Ser Asn Cys Ser Pro Thr Ser Gly Lys Gln Leu Asp Leu Leu Phe Ser
 645 650 655
 Val Thr Leu Thr Pro Arg Thr Val Asp Leu Thr Val Ile Leu Ile Ala
 660 665 670
 Ala Val Gly Gly Val Leu Leu Leu Ser Ala Leu Gly Leu Ile Ile
 675 680 685
 Cys Cys Val Lys Lys Lys Lys Lys Thr Asn Lys Gly Pro Ala Val
 690 695 700
 Gly Ile Tyr Asn Gly Asn Ile Asn Thr Glu Met Pro Arg Gln Pro Lys
 705 710 715 720
 Lys Phe Gln Lys Gly Arg Lys Asp Asn Asp Ser His Val Tyr Ala Val
 725 730 735
 Ile Glu Asp Thr Met Val Tyr Gly His Leu Leu Gln Asp Ser Ser Gly
 740 745 750
 Ser Phe Leu Gln Pro Glu Val Asp Thr Tyr Arg Pro Phe Gln Gly Thr
 755 760 765
 Met Gly Val Cys Pro Pro Ser Pro Pro Thr Ile Cys Ser Arg Ala Pro
 770 775 780
 Thr Ala Lys Leu Ala Thr Glu Glu Pro Pro Pro Arg Ser Pro Pro Glu
 785 790 795 800
 Ser Glu Ser Glu Pro Tyr Thr Phe Ser His Pro Asn Asn Gly Asp Val
 805 810 815
 Ser Ser Lys Asp Thr Asp Ile Pro Leu Leu Asn Thr Gln Glu Pro Met
 820 825 830

Glu Pro Ala Glu
835

<210> 5
<211> 23
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 5
accgcctcaa cttgttcaca tgg

23

<210> 6
<211> 26
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 6
ctgggtctcag gagccagcaa cttgtc

26

<210> 7
<211> 25
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 7
ctctatgacgt ggcagtttgt cgttc

25

<210> 8
<211> 26
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 8
ggctcgctca ttactcaagt caaccca

26

<210> 9
<211> 36
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 9
attcgcgact gatgatcgat tttttttt tttttt 36

<210> 10
<211> 20
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 10
attcgcgact gatgatcgat 20

<210> 11
<211> 20
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 11
gagatattag aattctactc 20

<210> 12
<211> 17
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 12
gagtagaatt ctaatat 17

<210> 13
<211> 22
<212> DNA
<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 13

agtccatgtg aacaagttga gg

22

<210> 14

<211> 20

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 14

aattctccca cctcagcctc

20

<210> 15

<211> 22

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 15

aggatgaaaa cgacaatgtg cc

22

<210> 16

<211> 21

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 16

agaattgctt gagcccagga g

21

<210> 17

<211> 21

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 17

caacttcaca ttgctcagtg g

21

<210> 18
<211> 25
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 18
tgagcaagg tt cagcctgg tt aagtc

25

<210> 19
<211> 26
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 19
caccgaaatac tcataaagaa ggtccc

26

<210> 20
<211> 26
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 20
tagacttcga gcaggagatg gccact

26

<210> 21
<211> 20
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 21
ccagccatgt acgttagccat

20

<210> 22
<211> 19
<212> DNA
<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 22

ccaagaagga aggctggaa

19

<210> 23

<211> 25

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 23

ccatcaccat cttccaggag cgaga

25

<210> 24

<211> 19

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 24

ccaagaagga aggctggaa

19

<210> 25

<211> 20

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 25

tgcaggaggc attgctgatg

20

<210> 26

<211> 19

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 26

aaatcgtgca cttgcaggc

19

<210> 27
<211> 18
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 27
ttgatgcgtt ccagctga

18

<210> 28
<211> 21
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 28
ttgaattcac tgtgtggagc c

21

<210> 29
<211> 19
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 29
tgcaggcaac agtgatgtc

19

<210> 30
<211> 24
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 30
atggcccttc cttaggctgg ctac

24

<210> 31
<211> 43
<212> DNA
<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 31

tgttagcgtga agacgacaga aagggcgtgg taccgagctc gag

43

<210> 32

<211> 22

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 32

agggcgtgg accgagctcg ag

22

<210> 33

<211> 11

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 33

ggctcgagct c

11

<210> 34

<211> 22

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 34

ggccatgtcc ggtgggcttg tg

22

<210> 35

<211> 26

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer

<400> 35

ctcaaaaactc ctggacaagt tgctgg

26

<210> 36
 <211> 22
 <212> DNA
 <213> Künstliche Sequenz

 <220>
 <223> Beschreibung der künstlichen Sequenz: Primer

 <400> 36
 aaggtgagg tcggagtcaa cg

22

<210> 37
 <211> 24
 <212> DNA
 <213> Künstliche Sequenz

 <220>
 <223> Beschreibung der künstlichen Sequenz: Primer

 <400> 37
 ggcagagatg atgacccttt tggc

24

<210> 38
 <211> 23
 <212> DNA
 <213> Künstliche Sequenz

 <220>
 <221> 5'UTR
 <222> (1)..(282)

 <220>
 <221> GC_signal
 <222> (147)..(157)

 <220>
 <221> misc_feature
 <222> (201)..(209)
 <223> cap signal; Transkriptionsstart

 <220>
 <221> 3'UTR
 <222> (2794)..(6163)

 <220>
 <221> 3'UTR
 <222> (2794)..(6163)

 <220>
 <221> CDS
 <222> (283)..(2793)

 <400> 38
 agcagcagaa ccccttagcag tgc

23

<210> 39
<211> 26
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 39
agaaccctta gcagtgcgat agagac 26

<210> 40
<211> 27
<212> DNA
<213> Künstliche Sequenz

<220>
<223> Beschreibung der künstlichen Sequenz: Primer

<400> 40
gaactgtaat gttgcttct cgtggca 27

Fig.1A

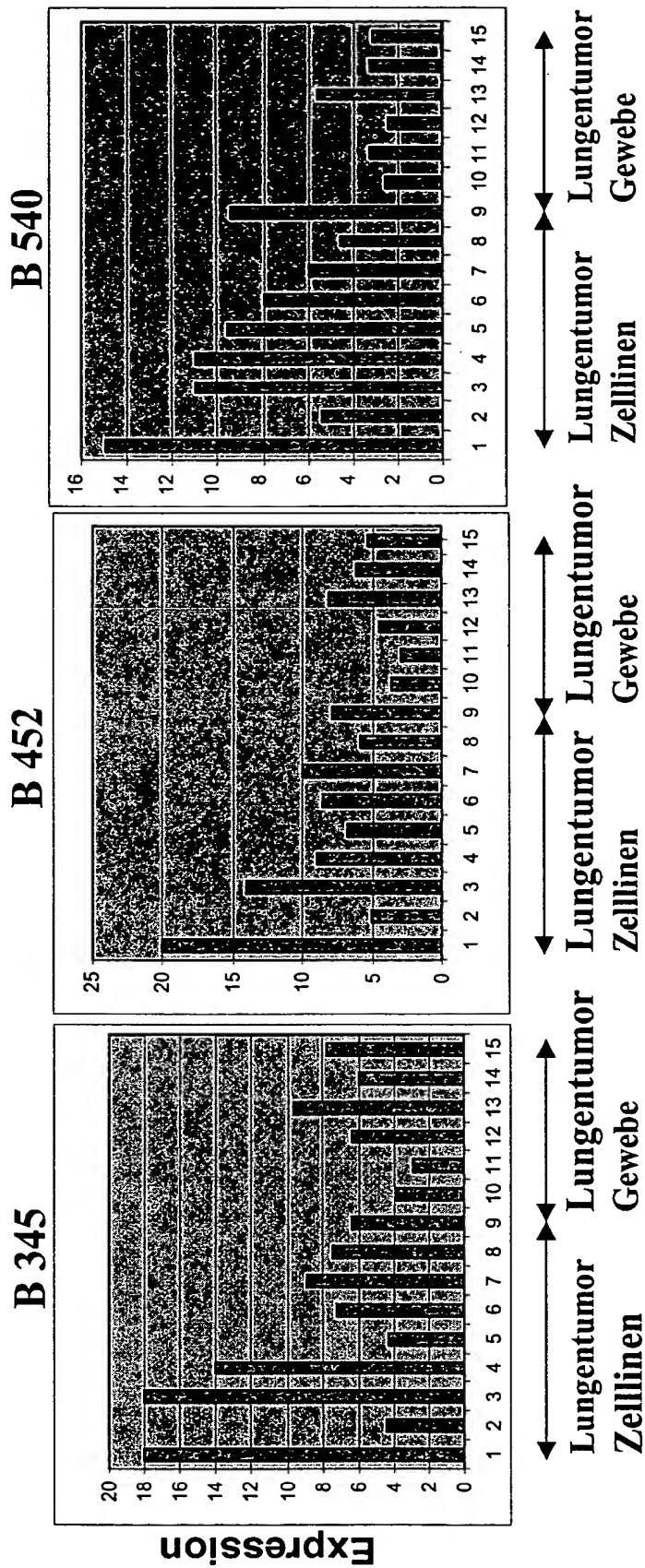


Fig. 1B

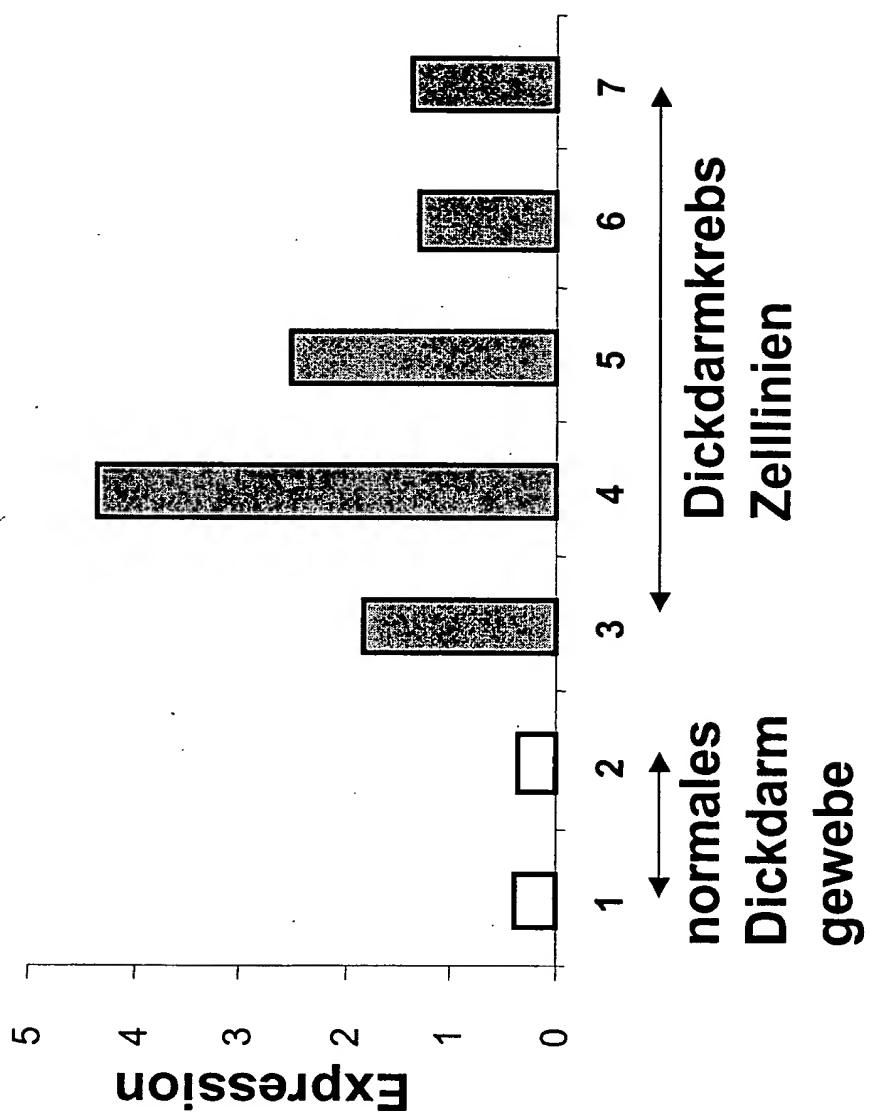


Fig. 1C

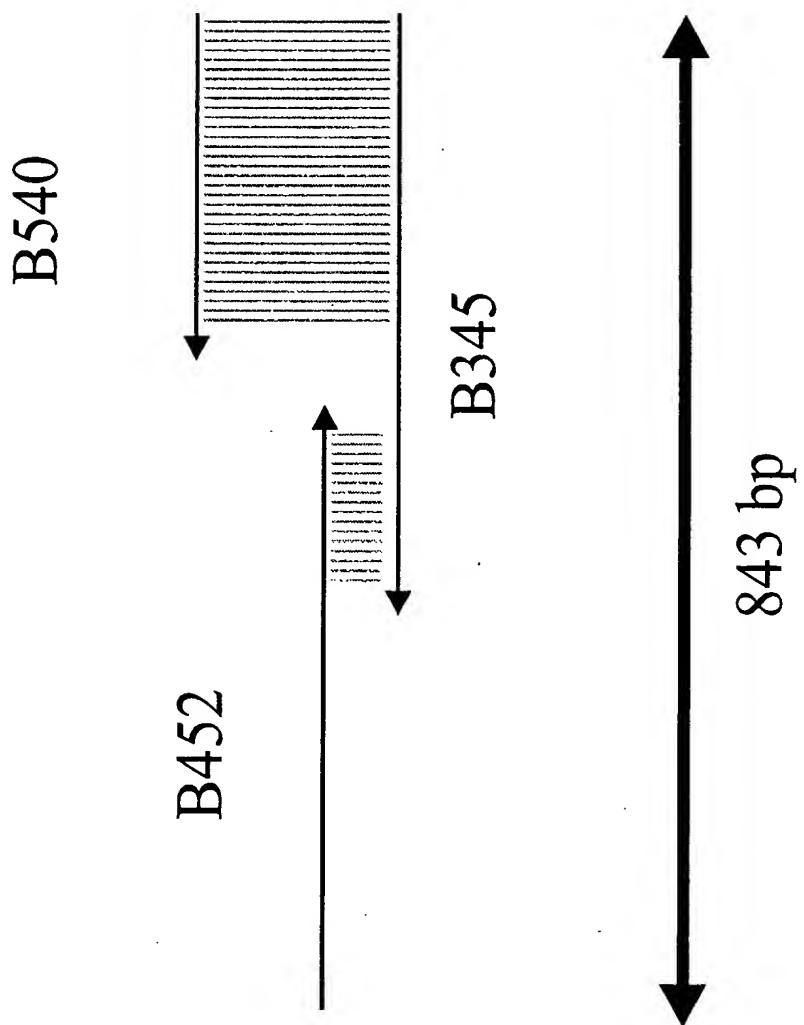


Fig. 2A

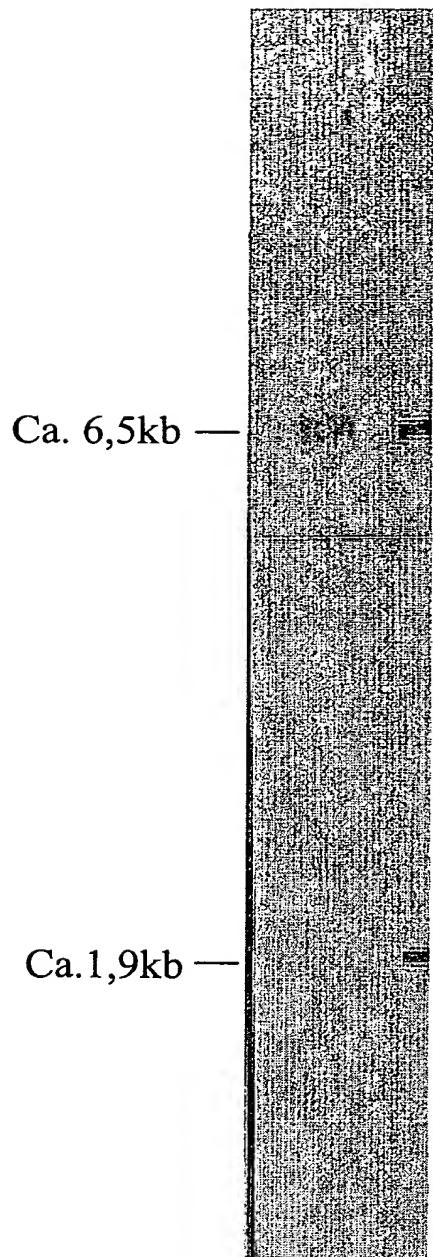


Fig. 2B

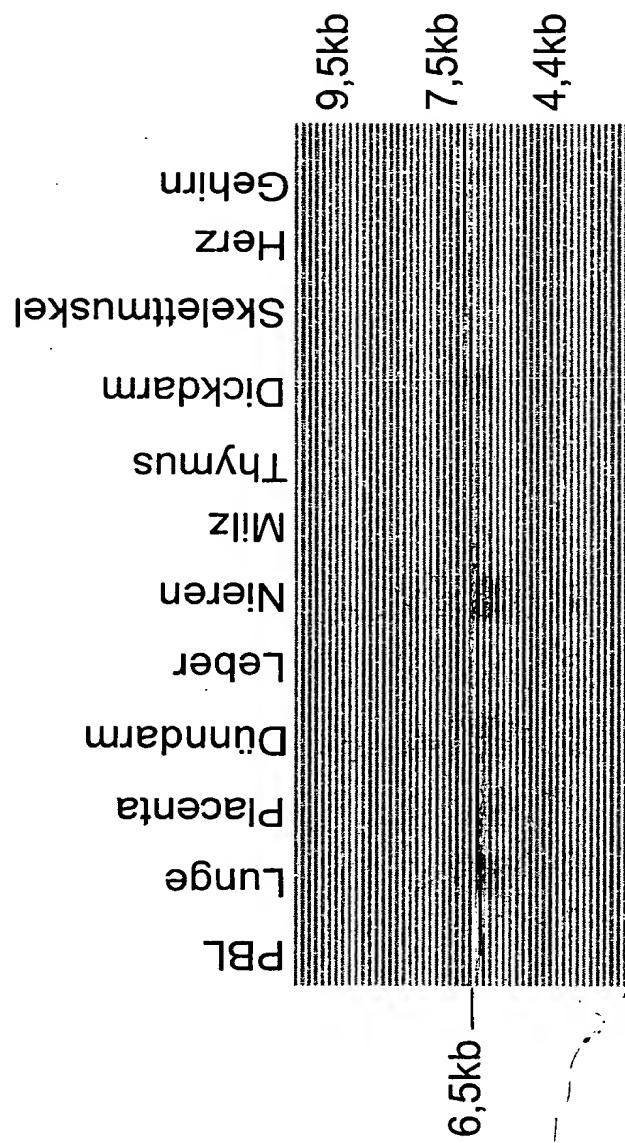


Fig. 2C



B345/AKTIN

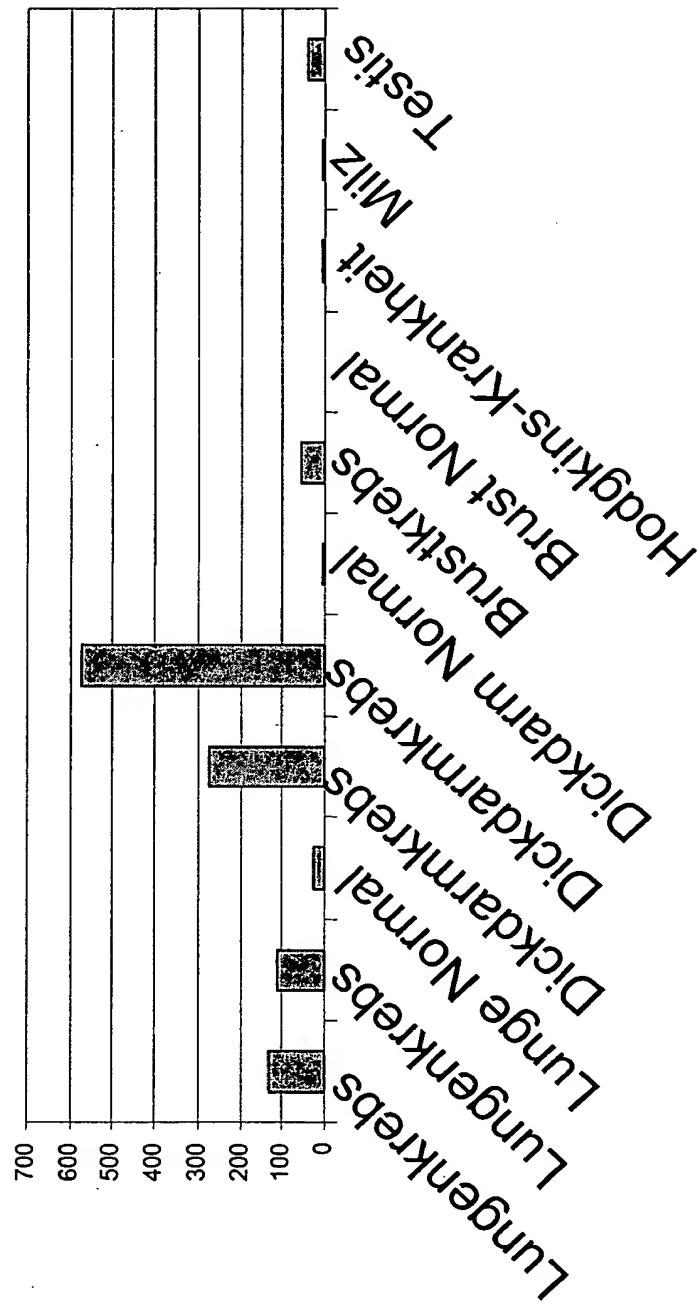


Fig. 3b

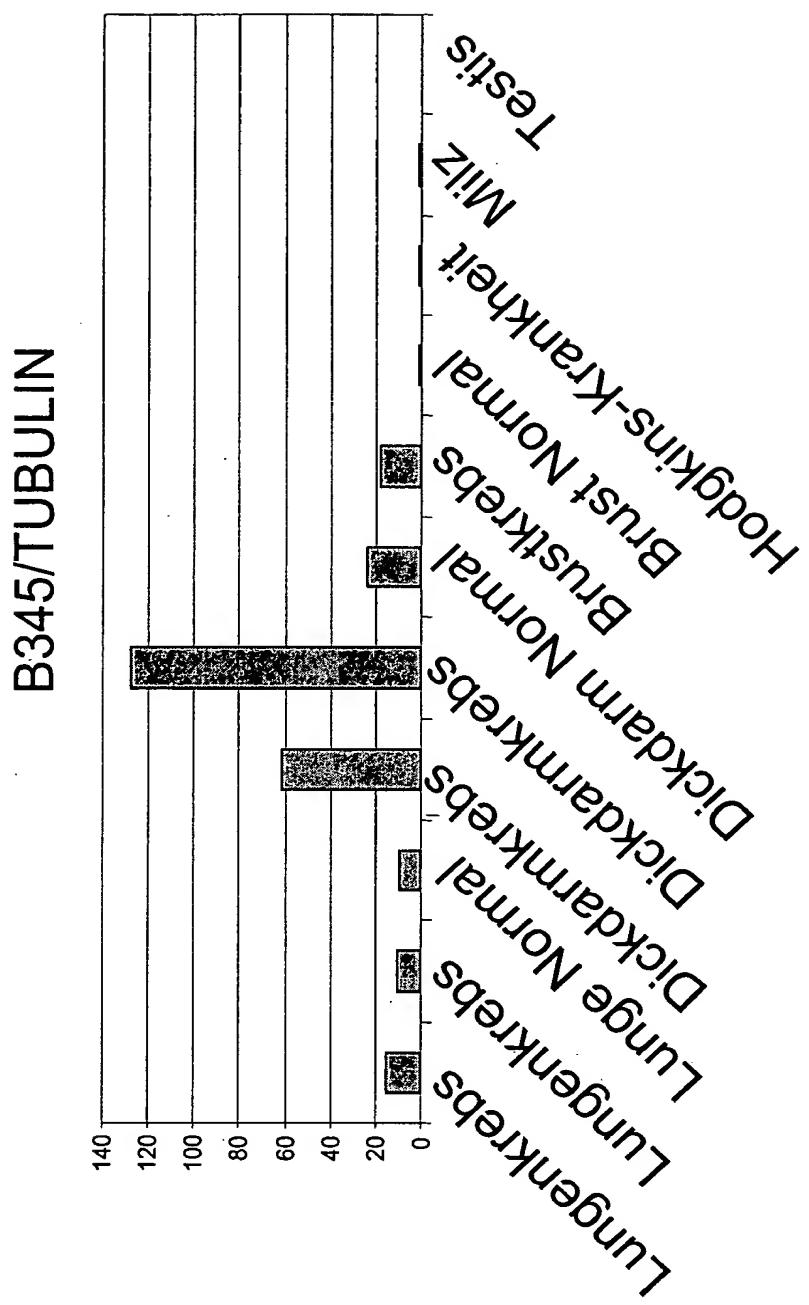


Fig. 4

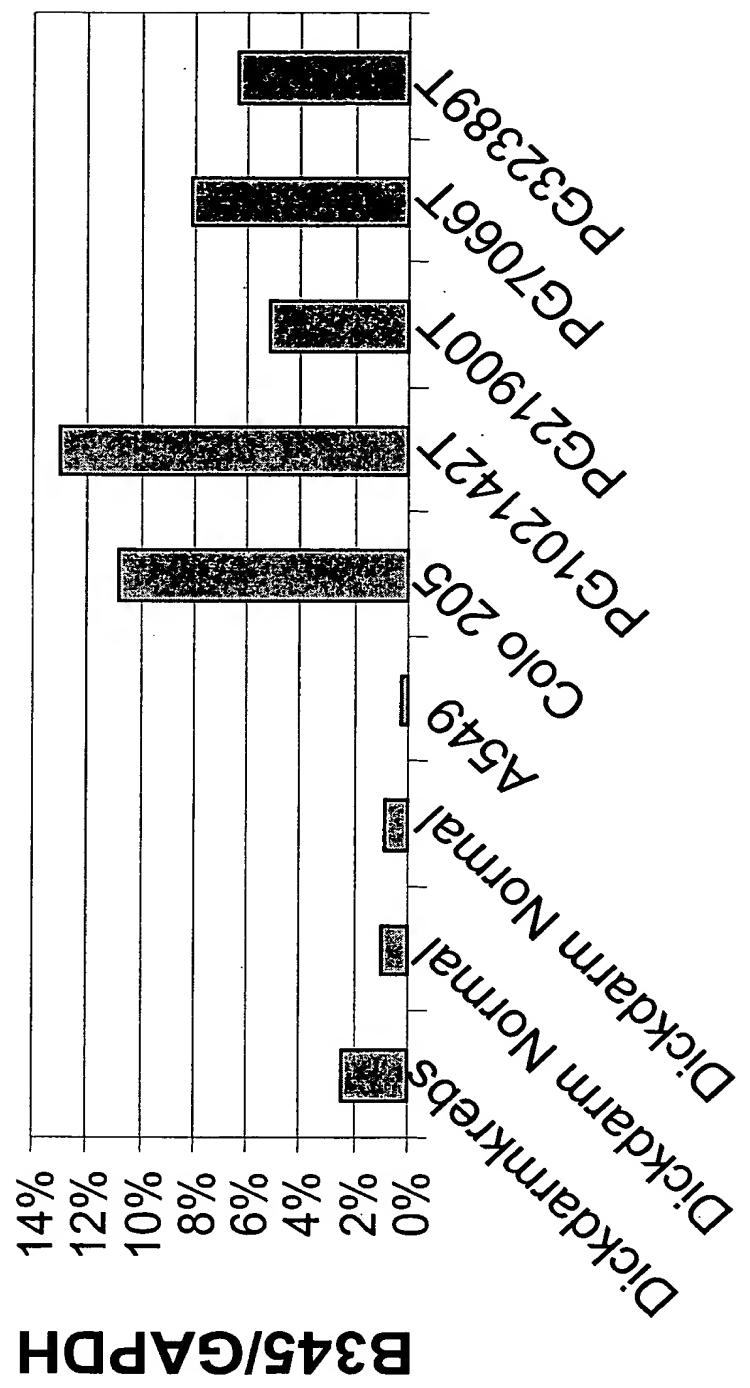


Fig. 5

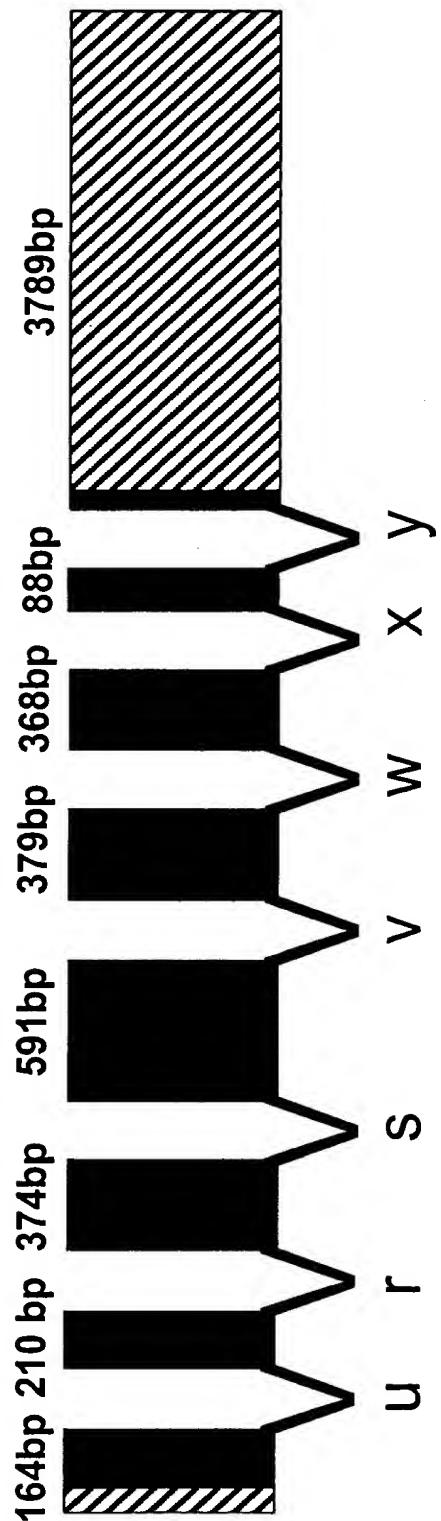


Fig. 6

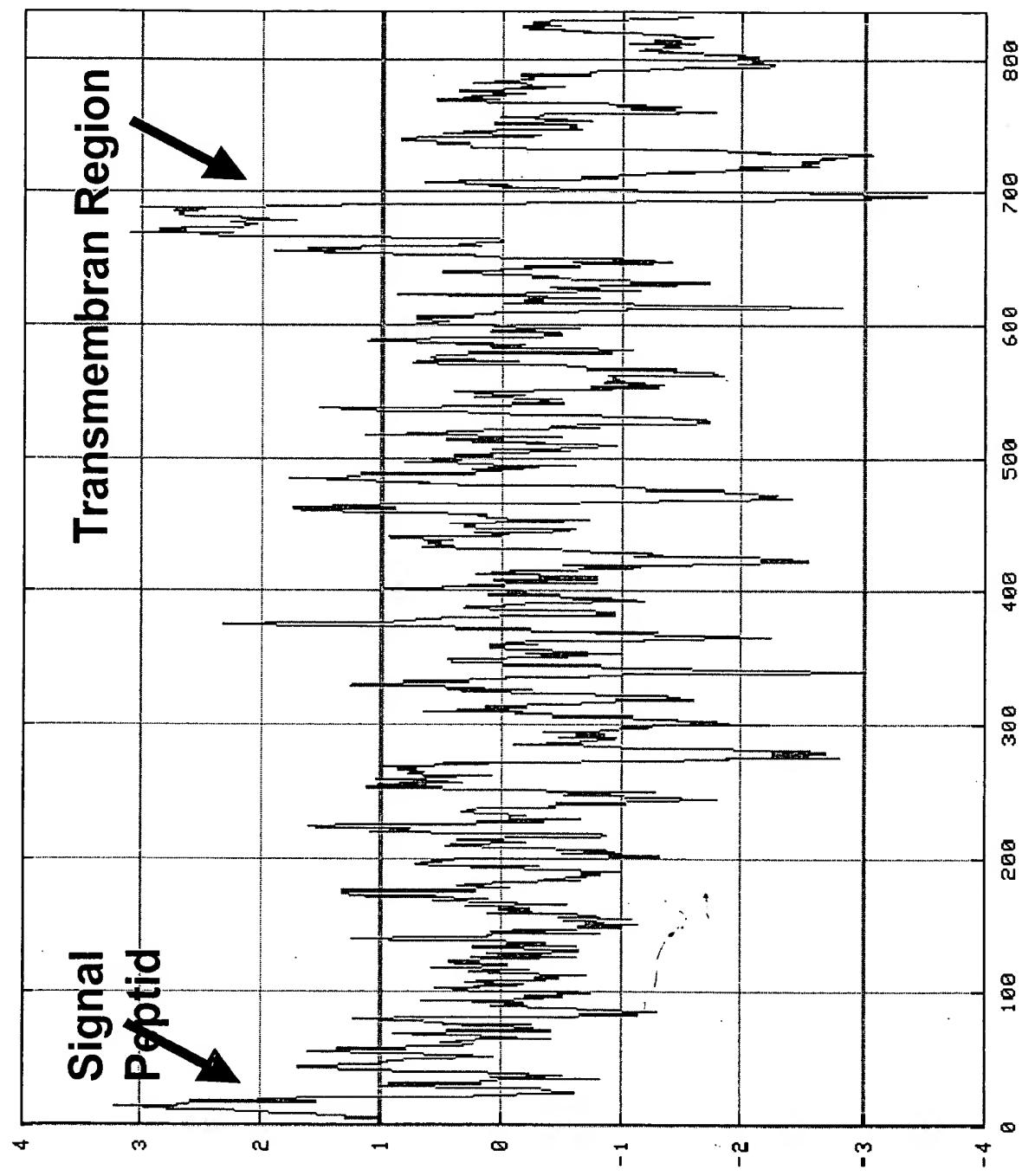


Fig.7

